31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

A Motion Sequence Fusion Technique Based on PCA for Activity Analysis in Body Sensor Networks

Hassan Ghassemzadeh, Eric Guenterberg, Sarah Ostadabbas, Roozbeh Jafari
Embedded Systems and Signal Processing Lab, Department of Electrical Engineering
University of Texas at Dallas, Richardson, TX, 75080
Email: {h.ghassemzadeh, mavpion, sarahostad, rjafari}@utdallas.edu

Abstract—**Human movement analysis by means of mobile sensory platforms is an ever-growing area with promise to revolutionize delivery of healthcare services. An effective data fusion technique is essential for understanding the inertial information obtained from distributed sensor nodes. In this paper, we develop a data fusion model based on the concept of principal component analysis. Unlike traditional fusion techniques which deal with statistical feature space, our model operates on motion transcripts, where each movement is represented as a sequence of basic building blocks called primitives. We describe how our model transforms transcripts of different nodes into a unified transcript by integrating the most relevant primitives of movements. Finally, we demonstrate the performance of our transcript fusion model for action recognition using real data collected from three subjects.**

I. Introduction

Basic human movement expresses a wealth of medically relevant data. Tremors and gait characteristics can be used to diagnose and rank several degenerative diseases such as Alzheimer's and Parkinson's. The specific activities a person performs during the day can be used to assess quality of life and fitness recommendations. In addition, activity analysis can be used for rehabilitation programs and sports training.

Video and image analysis from carefully placed cameras is a traditional platform for activity analysis [1]. This platform is limited by the inability to follow the patients beyond the reach of the cameras. Privacy is another major concern created by cameras. Both of these problems can be addressed by placing intelligent sensors directly on the patient. These sensors form a wireless network and process the data immediately, allowing the patient to be monitored pervasively. Sensor data is considered less invasive than video or audio data, making patients more willing to be monitored in this fashion. Typically, some combination of accelerometers, gyroscopes, and magnetometers are used because of their ability to precisely capture movements. This platform is referred to as a Body Sensor Network (BSN).

BSNs can generate a prodigious quantity of data. Unless this data is significantly reduced, with relevant sections automatically highlighted, it will be useless to healthcare providers and other users. In some cases, full automatic recognition is appropriate, as in life-logging, where the provider is given a summary or schedule of the activities of the patient. However, for some sections, a physician may wish to see more detailed analysis.

One interesting approach for representing data is something we call motion transcripts. Motion transcripts are generated automatically from sensors on a single limb by grouping similar readings. Each group is labelled with a symbol and represents a unique motion, called a motion primitive. A transcript is a sequence of labels corresponding to the sensor readings. Motion primitives often span several samples and may appear in non-contiguous portions of the transcript.

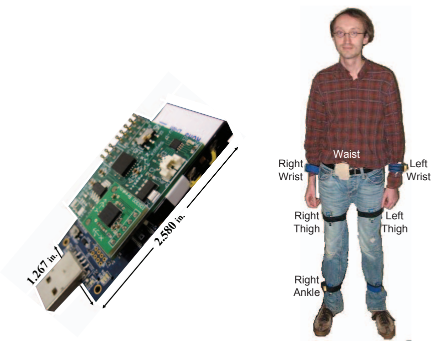
When transcripts from several sensor locations are generated, the data quickly becomes overwhelming. It is necessary to somehow combine, or fuse, all the individual transcripts to create a unified transcript called a choreography. Assigning a unique symbol to every unique combination of symbols from individual limbs results in overly detailed and confusing choreographies. Further, limb movements are not completely synchronized, so movements where one limb moves slightly faster or slower could result in a very different choreography for similar movements.

This paper addresses this problem with a technique for creating a simplified choreography by selecting only the most relevant symbols from the individual transcripts. This selection is performed using a technique based on Principal Component Analysis (PCA). In this paper we will describe the generation of motion transcripts and the fusion technique that creates a choreography. We show results from various transitional actions and validate the results both visually and with automatic action recognition based on the transcripts.

II. Related Works

Several researchers have considered the idea of representing complex movements with sequences of motion primitives. Much of this research is motivated by phonological analysis, which takes raw, continuous speech sounds and divides them into phonemes, which are further grouped into words. For instance, the authors of [2] show how to decompose joint angles calculated from visual data into a visuo-motor language called HAL (Human Activity Language). HAL forms primitives by segmenting data at the zero crossings of angular velocity and acceleration, then combines primitives to form a dictionary of activities. Another approach [3] uses a clustering

978-1-4244-3296-7/09/$25.00 ©2009 IEEE



(a) Sensor node (b) Sensor Placement

Fig. 1. Sensor Node and Placement on Experimental Subject

algorithm to organize action primitives. The authors of [4] describe how to modify the Isomap algorithm to handle motion data with both temporal and spatial components, which can enable the discovery of motion primitives. Previously, we built motion transcripts using Gaussian Mixture Models to reduce transmission energy by reducing the multi-dimensional per-sample observations on a sensor node to a single character taken from a small alphabet [5].

This work uses eigenvectors generated through PCA to pick representative motion primitives from individual transcripts. This is related to work by Cohen *et al* [6] in which a subset of features are chosen using the same criteria as the PCA. They call this method Principal Feature Analysis (PFA). Similarly, the authors in [7] consider, both theoretically and empirically, the topic of unsupervised feature selection using PCA, by leveraging algorithms for the so-called Column Subset Selection Problem (CSSP). In words, the CSSP seeks the best subset of exactly k columns from an $m \times n$ data matrix A, and has been extensively studied in the Numerical Linear Algebra community.

III. System Architecture

This research is based on a BSN, comprising several wireless sensor nodes placed on subjects, to monitor daily activities. The sensor nodes are commercially available TelosB motes from XBow®. We use the custom-designed sensor board shown in Fig. 1(a). Each sensor node includes a three-axis, 2g accelerometer and one two-axis gyroscope. The nodes sample at about 20 Hz and use a TDMA scheme to communicate all data to an off-body base station. The sampling rate has been experimentally chosen to provide sufficient resolution while compensating for the bandwidth constraints of our sensor platform. Six sensor nodes are placed on the subjects, as shown in Fig. 1(b). The base station relays the information to a PC via USB. For the purpose of the research, all further analysis is performed in MATLAB.

IV. Motion Transcript Fusion

As mentioned in the introduction, we combine individual transcripts of limb motion to produce a unified transcript of human motion called a choreography. Our data fusion model, shown in Fig. 2 aims to generate a choreography which incorporates the most significant information from all sensor nodes.

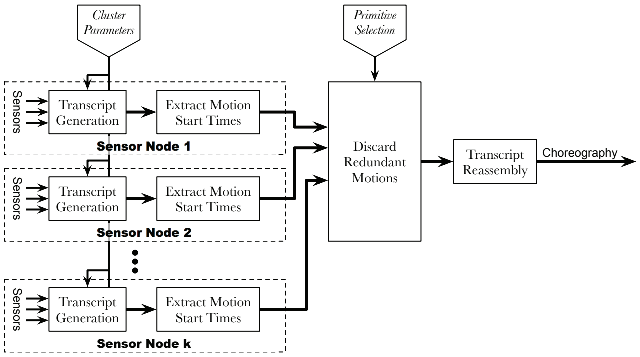


Fig. 2. Application of the Fusion Model

A. Motion Transcript

Physical movements can be represented by coordinated sequences of simple motions and postures. Each limb has its own sequence of motions that is coordinated with and affected by the motions on other limbs. For instance, during walking, the foot lifts off the ground, moves forward, touches the ground, and bears the weight of the body. These motions generate accelerations and rotations that can be recorded by on-body inertial sensors.

The goal of transcript generation is to use the sensor data to create a sequence of symbols, called motion primitives, corresponding to the limb motions. Important information about the motion at any given time may be contained in a short interval of sensor readings centered on the time of interest. These short, overlapping intervals are called frames and can be individually assigned symbols using classification techniques.

Statistical classification uses training data to create a model which can be used to assign labels to unlabelled data. If the frame labels for the training data are unknown, methods known as clustering can group frames together based on statistical features, such as mean, standard deviation, root mean square, first and second derivatives from a moving window centered about the current point. Individual data points are then labelled

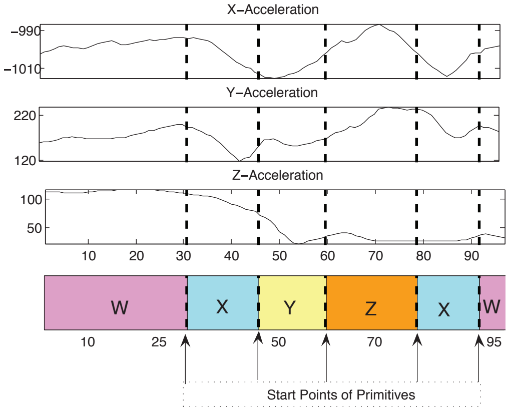


Fig. 3. Example Transcript

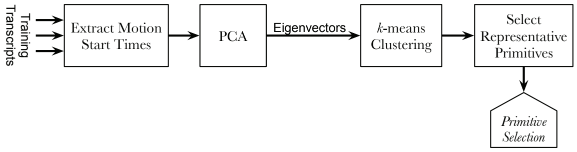


Fig. 4. Training the Fusion Model

based on the derived clusters. The centroid of each cluster defines a movement primitive.

Primitives are extracted from a set of training movements. Frames are assigned to the primitive whose centroid they are nearest to. Transcripts show the sequence and timing of primitives for each sensor node separately from the sequences for the others. Fig. 3 shows a sample transcript generated by the node placed on the *right wrist* for one trial of the "Bend and Grasp" movement. The transcript consists of four primitives labeled as "W", "X", "Y", and "Z", each illustrated by a separate color. The start times of the primitives are marked by arrows. For this specific trial, two of the primitives ("W" and "X") are repeated twice. Our fusion model uses such timing information about primitives to build transcript of the overall network.

We use a k-means clustering technique to define our primitives because it is algorithmically simple and efficient to use after training. A fundamental question when using k-means clustering is the proper value of k, which determines the number of clusters the algorithm will find. In our system, too few clusters will cause the transcript to miss key details that are crucial for evaluating movements, and too many clusters will produce irrelevant details that will prevent the system from generating consistent transcripts for similar movements. We currently choose k by trial and error, but we are investigating automated techniques for determining the optimal number of clusters.

B. Fusion Model

As discussed earlier, our fusion model takes transcripts of different nodes as input and generates a choreography representing overall body movement. To build such model, a set of training transcripts is used as shown in Fig. 4. Since each transcript is a sequence of basic motions in time, it can be represented by the start time of its primitives. For each of the training movements, timing data are extracted from transcripts of all sensor nodes. The choreography is generated from a subset of these start times. Selected times should contain the most significant information about each action. We use a PCA-based feature selection technique, called Principal Feature Analysis (PFA) [6], which takes a set of start times as input and produces the times that are best representative of the movements.

In our model, we use the eigenvector matrix produced by PCA to select prominent features. Assume A is the matrix whose columns are the orthogonal eigenvectors. Each row in the matrix is associated with one of the input start times. The key idea is that, rows with similar absolute values correspond to correlated inputs (timing information). We use k-means clustering on euclidean space to discover this similarity. Clustering groups primitives with similar timing information together. For each of the clusters generated by k-means one element is selected as representative of that group. The closest element to the mean of each cluster is the candidate for being representative. This way, a set of primitives which is most significant across the network is selected. This information will be used to assemble a choreography as shown in Fig. 2.

V. Action Recognition

Motion transcripts can be used in a variety of ways depending on the application of the BSN. In Section VI, we will test the effectiveness of our fusion model for movement classification. A number of classification algorithms have been used in the field of pattern recognition and machine learning. The k-Nearest-Neighbor (k-NN) method uses a similarity measure to find the action in the training set most similar to the test data point. We use this technique to identify actions. Using k-NN algorithm for action recognition, an unknown movement is classified by a majority vote of its neighbors, with the action being assigned to the movement most common amongst its k nearest neighbors. If $k = 1$, the action is simply assigned to the class of its nearest neighbor.

For a classification algorithm to identify an unlabelled trial, a measure of similarity between different observations is required. Euclidean distance is widely used as the similarity measure when the training set is constructed based on statistical features. In our system, however, each movement is represented by a set of transcripts. Therefore, a similarity metric is required to find the difference between two strings. The *Levenshtein distance* [8], also called *edit distance*, is a well-known metric for measuring the amount of difference between two character sequences. The edit distance between two strings is given by the minimum number of operations needed to transform one string into the other, where an operation is defined as an insertion, deletion, or substitution of a single character.

Given two strings T and T', there are three basic operations to convert one to another. These operations include substitution (ith symbol of T is replaced by jth symbol of T'), insertion (jth symbol of T' is inserted into ith position in T) and deletion (ith symbol of T is deleted). We use the classic dynamic programming solution with a modified version of edit distance to calculate the amount of dissimilarity between two given fused transcripts. The cost of substitution is 1 except when replacing a null character, which incurs no cost. Null characters can appear in a choreography if no selected movement primitive covers a specific period of time.

VI. Experimental Results

In this section, we demonstrate the effectiveness of our fusion model for action recognition where movements of interest are represented by transcripts. The purpose of action recognition is to classify unknown transitional movements according to a set of pre-specified training actions. We arranged our

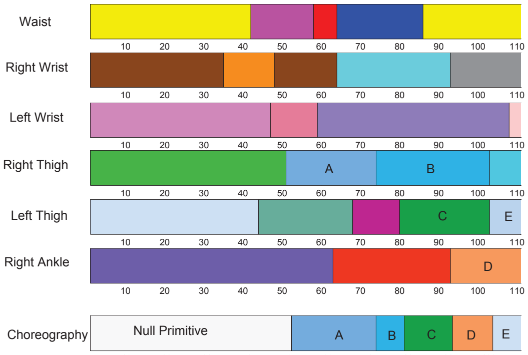


Fig. 5. Transcripts of individual sensor nodes and final transcript generated based on PCA for one trial of movement "Look back clockwise"

experiments to detect the ten movements: "stand to sit", "sit to lie", "bend and grasp", "kneeling", "turn clockwise", "look back clockwise", "step forward", "step sideways", "reach up to the cabinet", and "jumping". The six sensor nodes shown in Fig. 1(b) were placed on three healthy subjects to capture motions of upper and lower body joints during each movement. Subjects were asked to perform each movement for ten times. Five of these trials were used for model generation and training. The rest of the data were used for validation.

Motion transcripts were generated at each node by extracting several statistical features from each sample point and using k-means to group similar points together. Simple features such as *mean value*, *standard deviation*, *RMS power*, and *first* and *second derivative* formed the feature space for clustering. The transcripts of the six nodes were then used by our fusion model to generate a unique choreography for each trial. Fig. 5 shows sample transcripts generated by different nodes for one trial of "look back clockwise". Each color represents one specific primitive. As mentioned earlier, sensor nodes generate their transcripts independently. The last row shows the choreography generated by the PCA-based fusion model. For this specific example, five significant primitives were detected by the model which are labelled as "A", "B", "C", "D", "E" in the figure. The null symbol represents the region for which no prominent primitives were detected.

Choreographies of the training trials were used to train a k-NN classifier which operates on motion strings. This model was then used to classify each of the test trials based on its edit distance to the training transcripts. The overall accuracy was 85.33%. To compare this result with the performance of the system without transcript fusion, we built a new transcript for each motion trial. This transcript was formed by concatenating transcripts of the same trial from all the nodes. This way, all timing information was preserved within the new transcript. The resulting transcripts went through the same classification procedure as fused transcripts, achieving an accuracy of 91.00%. Fig. 6 shows per-movement classification accuracy for the concatenated transcripts and the choreography created with our technique. In some cases, simplicity may be preferred over accuracy. Our fusion technique reduces dimensionality of original data into a single transcript. It further simplifies representation of movements while maintaining fairly high classification accuracy.

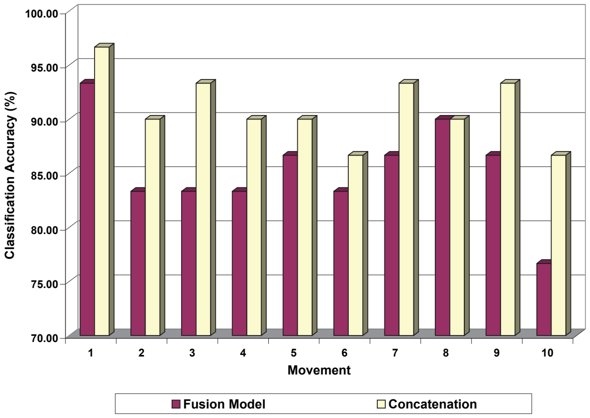


Fig. 6. Classification Accuracy Using Transcripts

VII. Conclusion

In this paper, we presented a PCA-based technique for fusion of motion sequences. First, motion transcripts are generated by grouping sample points with consistent physical behavior together. Each group is called a primitive and its timing information is fed into PCA for feature selection. Prominent primitives reported by PCA are then used to generate a unique transcript which is the best representative of all sensor nodes. We tested the performance of our model for action recognition. Through our experimental analysis on real data, we showed that our technique can achieve reasonably high classification accuracy.

Acknowledgment

The authors would like to thank Ville-Pekka Seppa for his efforts in designing the sensor board.

References

[1] A. Karunanidhi, D. Doermann, N. Parekh, and V. Rautio, "Video analysis applications for pervasive environments," in *Proc. 1 st International Conference on Mobile and Ubiquitous Multimedia, Oulu, Finland*, 2002, pp. 48–55.

[2] G. Guerra-Filho, C. Fermuller, and Y. Aloimonos, "Discovering a language for human activity," in *Proceedings of the AAAI 2005 Fall Symposium on Anticipatory Cognitive Embodied Systems, Washington, DC*, 2005.

[3] Z. Husz, A. Wallace, and P. Green, "Human activity recognition with action primitives," in *IEEE Conference on Advanced Video and Signal Based Surveillance, 2007. AVSS 2007*, 2007, pp. 330–335.

[4] O. Jenkins and M. Mataric, "Automated derivation of behavior vocabularies for autonomous humanoid motion," in *Proceedings of the second international joint conference on Autonomous agents and multiagent systems*. ACM New York, NY, USA, 2003, pp. 225–232.

[5] E. Guenterberg, H. Ghasemzadeh, V. Loseu, and R. Jafari, "Distributed continuous action recognition using a hidden markov model in body sensor networks," in *DCOSS*, ser. Lecture Notes in Computer Science, B. Krishnamachari, S. Suri, W. R. Heinzelman, and U. Mitra, Eds., vol. 5516. Springer, 2009, pp. 145–158.

[6] I. Cohen, Q. Tian, X. Zhou, and T. Huang, "Feature selection using principal feature analysis," *Urbana*, vol. 51, p. 61801.

[7] C. Boutsidis, M. Mahoney, and P. Drineas, "Unsupervised feature selection for principal components analysis," 2008.

[8] V. Levenshtein, "Binary Codes Capable of Correcting Deletions, Insertions and Reversals," in *Soviet Physics Doklady*, vol. 10, 1966, p. 707.