Kinect based People Identification System using Fusion of Clustering and Classification

Aniruddha Sinha[1], Diptesh Das[1], Kingshuk Chakravarty[1], Amit Konar[2] and Sudeepto Dutta[3]
[1]*Innovation Lab, Tata Consultancy Services Ltd., Kolkata, India*
[2]*Dept. of Electronics and Tele-communication Engineering, Jadavpur University, Kolkata, India*
[3]*Sikkim Manipal Institute of Technology, Sikkim, India*

Keywords: Kinect Sensor, Human Identification, Gait Detection, Clustering, Classification, Fusion, Dempster-Shafer Theory, Human Skeleton.

Abstract: The demand of human identification in a non-intrusive manner has risen increasingly in recent years. Several works have already been done in this context using gait-cycle detection from human skeleton data using Microsoft Kinect as a data capture sensor. In this paper we have proposed a novel method for automatic human identification in real time using the fusion of both supervised and unsupervised learning on gait-based features in an efficient way using Dempster-Shafer (DS) theory. Performance comparison of the proposed fusion based algorithm is done with that of the standard supervised or unsupervised algorithm and it needs to be mentioned that the proposed algorithm is able to achieve 71% recognition accuracy.

1 INTRODUCTION

Several modalities of human identification in terms of human computer interaction (HCI) exists in the current literature on machine intelligence. A few of the common techniques include voice, facial expression, gesture, iris, fingerprint etc. Unfortunately, all of these modalities demand direct human interaction and thus human identification in a non-intrusive manner is not feasible. Moreover extraction of fingerprint, iris or audio related biometric information (at recognizable form) from a large distance is also a challenging job. This paper aims at developing a novel scheme for human identification from the movement data pattern (gait) of the subject. The main advantages of gait based person identification is that it is unobtrusive and can be applied from a relatively large distance. And also the use of gait signature is highly secure because it is very much hard to hide and immitate. One approach to extract movement data is to determine the junction co-ordinates and their velocity to classify subject from the movement patterns. This however, requires multiple positional sensors or cameras to determine the co-ordinates and depth of junctions from a given distance. Fortunately, Microsoft Kinect provides an interesting platform to capture the junctional information using RGB camera and depth sensor in indoor environment. Not only that, as Kinect directly provides skeleton or junction information to store the movement pattern instead of video, this approach ensures the privacy and security issues for any individual.

Usually, classification (supervised learning) or clustering (unsupervised learning) algorithm is required to map or group the sensor captured data into classes or clusters. In case of supervised learning, the set of features extracted from the sensor data is often used to train a classifier for subsequent applications in classifying an unknown movement pattern. On the other hand, for unsupervised learning, the extracted features are used to recognize a new person without any training or manual labeling but grouping based on data similarity. It needs to be mentioned that these supervised and unsupervised algorithm based approaches for human identification had already been proposed by Adrian et al. (Ball et al., 2012) and Preis et al. (Preis et al., 2012) with different set of features. While Adrian et al. (Ball et al., 2012) only considered dynamic angular information extracted from movement pattern, Preis et al. (Preis et al., 2012) used both static information like height, length of torso etc. and dynamic information like step length and velocity. Hybrid features related to area of upper and lower body, and distance between the upper body centroid and the centriods derived from different body joints are proposed in (Sinha et al.,

Sinha A., Das D., Chakravarty K., Konar A. and Dutta S..
Kinect based People Identification System using Fusion of Clustering and Classification.
DOI: 10.5220/0004690201710179
In *Proceedings of the 9th International Conference on Computer Vision Theory and Applications* (VISAPP-2014), pages 171-179
ISBN: 978-989-758-009-3

2013). Artificial Neural Network (ANN) based connectionist framework for classification using 46 dimensional feature vector was demonstrated by Pal et al. (Pal and Chintalapudi, 1997). But, as the skeleton data provided by the Kinect is very noisy, a robust person identification algorithm is required to improve the recognition accuracy.

The supervised learning algorithm can be negatively influenced by limited quality of training data (Karem et al., 2012) as well as the parameters of the learning algorithms. Moreover it often results in misclassification due to asymmetrical distribution of real classes in training dataset. As an approach to overcome these shortcomings, this paper attempts to design a novel approach of human identification by fusing the score of both supervised and unsupervised algorithm. In addition, the proposed approach is very much insensitive to small variation in measurements due to spurious noise pick-ups, thus improves recognition accuracy.

The personnel identification method introduced above includes three main steps i) Acquisition of measurement related to movement data of subjects using Kinect sensor , ii) Feature extraction from the recorded skeleton data, iii) Decision making from extracted features. The feature extraction module extracts the half gait cycle automatically from the recorded movement data and compute all the features F mentioned in (Sinha et al., 2013). The decision making module referred above has three basic modules in its functional architecture i) a classifier **C** ii) a clustering algorithm **A** and iii) a fusion algorithm **D**. The classifier **C** is trained with extracted features set **F** of training dataset and tested with unknown input (movement) pattern. A classification measure in terms of probability $\mathbf{M_C}$ is calculated from the classification output of the test subject. The unsupervised clustering algorithm **A** is used to group the extracted features into three groups and a probabilistic measure $\mathbf{M_A}$ is computed according to the clustering result for the unknown test data. Then a fusion algorithm **D** is used to fuse the two performance measures $\mathbf{M_C}$ and $\mathbf{M_A}$ in order to improve the recognition accuracy of the unknown input. After a careful study, we have realized **C** by Support Vector Machine (SVM) (Cortes and Vapnik, 1995) with radial basis function (RBF) kernel and **A** by Fuzzy C-Means (Bezdek, 1981) (Bezdek et al., 1984) clustering. It is experimentally found that the combined SVM-FCM approach improves the recognition accuracy in comparison to the existing singleton classification or clustering algorithm mentioned in (Ball et al., 2012) (Preis et al., 2012) (Sinha et al., 2013). Any fusion algorithm like Bayesian networks (Dempster, 1968) (Jeffreys, 1973), Kalman filter (Kalman, 1960), Dempster-Shafer (Dempster, 1967) (Fine, 1977) can be used in the present scheme. We have used Dempster-Shafer algorithm for realizing **D** because of the following reasons.

- It is simple and generalizes probabilistic modeling and inference;
- It has the capability of making proper distinction between reasoning and decision taking;
- It properly quantifies
  1. The presence of confirming or contradicting information sources (termed as conflict).
  2. Low confidence and high confidence results (termed as ignorance) depending on availability of information sources.

This novel approach of fusing SVM & FCM output using Dempster-Shafer algorithm provides an outstanding performance on human identification from raw Kinect data in comparison to the existing research outcome in this arena (Ball et al., 2012) (Preis et al., 2012) (Sinha et al., 2013).

Rest of the paper is organized as follows. In Section II we explained the existing state-of-the-art methods in this context. Section III demonstrates the details of the proposed fusion based algorithm for person identification. The performance evaluation supported by experimental results are presented in Section IV. And finally we have concluded in Section V.

2 RELATED WORK

People identification using gait biometric has created a great interest in scientific community due to its non-intrusive nature of identification (Cheng et al., 2012). Gait recognition can be broadly classified into two categories 1) Model-based approach and 2) Model-Free approach (Carlsson, 2000) (Huang et al., 1999) . In the model-based approach gait signature are generated by modeling and tracking different body parts (like limbs, arms, thighs etc.) over time (Wang et al., 2010), (BenAbdelkader et al., 2002) where as in model-free approach, features are generated based on the change in shape of human silhouettes over time depending on the motion dynamics (Sarkar et al., 2005). Either of model-based and model-free approaches resolve the issues of each other but suffer from its own limitations. In model-based approach, the feature generation is view-invariant and scale-independent but sensitive to the quality of gait sequences and computationally expensive. Similarly in model-free approach the feature generation is insensitive to the quality of silhouettes and offers low computational overhead, but dependent on the viewpoint and scale. Direct gait classification using standard classifiers such as SVM (Support Vector Machine) (Cortes and Vapnik, 1995), K-NN (K-Nearest Neighbors) (Bouchrika and Nixon, 2008), TSVM (Transductive Support Vector Machine) (Dadashi et al., 2009), Dynamic Time Wrapping (DTW) (Kale et al., 2003) and HMM (Hidden Markov Model) (Cheng et al., 2008) (Meyer et al., 1998) are adopted by researchers as mentioned in (Ball et al., 2012) (Preis et al., 2012) (Sinha et al., 2013). Xeu et al. (Xue et al., 2010) proposed two fold fusion mechanism namely multiple feature fusion and multiple view fusion for human gait identification using Independent Component Analysis (ICA) and Genetic Fuzzy Support Vector Machine (GFSVM) (Xue et al., 2010) classifiers. Moreover, Dempster-Shafer algorithm had been used in (Le Hegarat-Mascle et al., 1997) for unsupervised classification in multi-source remote sensing where data fusion is done on pixel level to successfully identify landcover types. In this paper we have proposed a probabilistic approach of fusing the SVM & FCM output using Dempster-Shafer algorithm. This approach of fusing supervised and unsupervised learning accuracy in real-time human identification problem is novel to the best of author's knowledge and helps in improving the overall recognition accuracy.

3 ALGORITHMIC APPROACH

The proposed algorithm for people identification is based on the fusion of supervised and unsupervised algorithms. In this paper, Support Vector Machine (SVM) (Begg et al., 2005) (Cortes and Vapnik, 1995) with Radial Basis function kernel is used as supervised learning algorithm and for unsupervised learning we have used Fuzzy C-Means clustering algorithm (Bezdek, 1981)(Bezdek et al., 1984).The motivation is to partition the training data into three clusters with an aim to partition the subjects by height (small, medium, tall), by speed (slow, medium, fast), by the area encompassed (less, medium, large) or by the angles created by different segments of the body parts (low, medium, high). The centroids of these clusters would be stored during the training phase. Later during the testing, the distances of the new set of features would be computed from the centroids of the clusters to derive the probability of the test subject to belong to a particular cluster.This probability would be then fused using Dempster-Shafer (DS) theory (Dempster, 1967) (Fine, 1977) with the probability obtained from SVM during the testing phase. The uncertainty in the subjects in the clustering provides additional information to the fusion algorithm and we hope that this would improve the overall performance of the accuracy of detecting an individual.

3.1 Feature Extraction

The feature extraction is an important step in any machine learning based algorithm. In our personnel identification problem, the feature extraction is performed on every half Gait Cycle. A Gait Cycle is starting with one foot forward and ending with same foot forward. We have used a set of area (Sinha et al., 2013), static (Preis et al., 2012) and dynamic (Sinha et al., 2013) distances, certain angles (Ball et al., 2012) and the speed (Preis et al., 2012) as features derived from the 20 skeleton points. A summary of these features are given below.

Area Features (f_A).- Mean area occupied by upper (f_{au}) and lower (f_{al}) part of the body in half-gait cycle is $f_A = \{f_{au}, f_{al}\}$. The joints considered are given below:

Upper body - Shoulder centre, shoulder left, hip left, hip centre, hip right and shoulder right.

Lower body - Hip centre, hip right, knee right, ankle right, ankle left, knee left and hip left.

Distance Features (f_D). - The Euclidean distances of the adjacent joints are the static distances. The Euclidean distances of centroids of upper and lower limbs from the upper body centroid are the dynamic distances as they change while one is walking. The joints considered for computing the centroids are given below:

Upper body - shoulder centre, shoulder left, hip left, hip right and shoulder right.

Right hand - shoulder right, elbow right and wrist right.

Left hand - shoulder left, elbow left and wrist left.

Right leg - hip right, knee right and ankle right.

Left leg - hip left, knee left and ankle left.

Figure 1 shows the Euclidean distance between upper body centroid and right hand centroid.

Angle Features (f_G). - These are the angles that various segments of the two legs make with the horizontal and vertical planes (Ball et al., 2012).

Other Features (f_S). - Apart from the above mentioned features, we have also considered all the static and dynamic features mentioned in (Preis et al., 2012).

The combined gait feature vector used for people identification is of dimension

$$\vec{F} = \{f_A, f_D, f_G, f_S\} \in \mathbf{R^{46}} \quad (1)$$

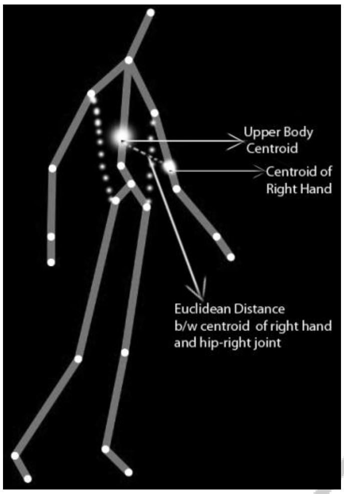

Figure 1: Euclidean distance between upper body centroid and right hand centroid.

3.2 Dempster-Shafer Theory (DST)

Dempster-Shafer Theory (Dempster, 1967) (Fine, 1977) , based on the principle of mathematical theory of evidence (Shafer, 1976), is one of the promising algorithms for combining evidence of different sources to arrive at a degree (generally termed as belief, defined by a belief function) to which all the current evidence supports . The theory was first proposed by Arthur P. Dempster (Dempster, 1967) and Glenn Shafer (Fine, 1977) (Shafer, 1976). It is mainly a generalized form of Bayesian theory of subjective probability (Dempster, 1968). Then naturally we will have to understand basics of Bayesian theory (Jeffreys, 1973) of subjective probability to apply the DST in our application.
If unconditional probability of an event A is denoted by P(A) and A has a domain of possible values $\{x_1, x_2, x_3, ..., x_n\}$, then the sum of all probabilities of A = x_1, A = x_2....A = x_n is always 1. In mathematical notation it can be expressed as

$$\sum_{i=1}^{n} P(A = x_i) = 1 \tag{2}$$

The conditional probability of A, given that the event B has already been occurred, denoted by P(A|B) is defined by

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{3}$$

Suppose we are given P(A) and the conditional probability P(A|B), then Bayes' law (Jeffreys, 1973) of conditional probability is defined as

$$P(B|A) = \frac{P(A|B) * P(B)}{P(A)} \tag{4}$$

If A has n subsets $A_1, A_2, A_3, ..., A_n$, such that A = $\cup A_i$, where i $\in$ [1,n] and B has m partitions $B_1, B_2, B_3, ..., B_m$, such that B = $\cup B_j$, where j $\in$ [1,m], then by Baye's law, P($B_j|A_i$) can be expressed as

$$P(B_j|A_i) = \frac{P(A_i|B_j) * P(B_j)}{\sum\limits_{k=1}^{m} P(A_i|B_j) * P(B_k)} \tag{5}$$

Bayesian theory mainly assigns a positive belief to each of the propositions but it does not consider disbelief of the proposition. DST addresses this problem by ensuring fusion of information by considering both their belief and disbelief. In DST, the set of all possible outcomes of a random experiment often referred as frame of discernment (FOD). If a random experiment has $\{x_1, x_2, ..., x_n\}$ outcomes then FOD is defined as the universal set.
$\theta = \{x_1, x_2, ..., x_n\}$, where cardinality of θ is n = |θ| and 2^n subsets of θ are called propositions. In DS theory, we used to assign probability masses to the subsets of θ unlike Bayesian theory where each element is treated as singleton subject. When a source of evidence assigns probability mass to one of the propositions, the resulting function is termed as basic probability of assignment (BPA). In formal notation BPA is m

$$\begin{aligned}
&where\ \ m : 2^{\theta} \rightarrow [0,1] \\
&where\ \ 0 \leq m(x_i) \leq 1\ \ and \\
&m(\varphi) = 0\ (\varphi\ is\ empty\ subset\ of\ \theta) \\
&and \sum_{x \subseteq \theta} m(x) = 1.0
\end{aligned} \tag{6}$$

Belief function for DST over FOD is expressed as

$$Bel(X) = \sum_{Y \subseteq X} m(Y) \tag{7}$$

Similarly Uncertainty U(X) of DST is the measure to which we consider nothing one way or other about proposition. Plausibility for DST is defined as
Pl(X) = Bel(X) + U(X)
In this theory, probability of a set A $\in 2^{\theta}$ is bounded by Bel(A) $\leq$ P(A) $\leq$ Pl(A). Dempster's rule of combination (Jsang and Pope, 2012) is used to combine two independent set of probability masses. If there are two FOD θ1 and θ2 submitted by two source of information K1 and K2 respectively and BPA for θ1 and θ2 are $m_1(.)$ and $m_2(.)$, then the combination of probability masses (called joint mass) is computed in the following manner

$$\begin{aligned}
m_{1,2}(X) &= K^{-1} \sum_{X = X_i \cap X_j \neq \varphi} m_1(X_i) * m_2(X_j) \\
and\ \ K &= 1 - \sum_{X_i \cap X_j = \varphi} m_1(X_i) * m_2(X_j)
\end{aligned} \tag{8}$$

where X_i and X_j are focal elements of θ1 and θ2. The equation (8) is also called orthogonal summation of belief of functions.

3.3 Application of Dempster-Shafer Theory (DST) in Person Identification

In our proposed approach, we have combined decision of clustering and classification algorithm using DST to improve the recognition accuracy.

At first we have used supervised learning approach to solve the problem. In our experiment we have realized supervised learning algorithm by Support Vector Machine (SVM). Let us consider a case, where N-person skeleton data is available for learning and an unknown input pattern needs to be classified. For this, we have extracted P-dimensional feature vector for each of the N persons and form a training dataset D. Then we can represent each person by a M×P matrix $X_i \; \forall \; i \in [1,N]$) where rows ($D_{dim}$) of X_i denotes observation and column represents features. Here M is the total number of rows i.e. observations for the i^{th} person. The dataset D can be expressed in terms of X_i i.e D = $\{X_1, X_2, ..X_i, ..X_N\}$ (D is of dimension $D_{dim} \times P$). Now supervised learning algorithm SVM is used to generate a training model using the data set D. The testing is done using this training model or in the other words for an unknown input of dimension $K \times P$, SVM use the training model to label each of the K observations of the same. Once the labeling is done for each observation, we generate a probability score based on the distribution of labeling for the unknown input among all the N classes. If n_i $(n_i <= K)$ be the number of observations detected as class i, then the probability score of unknown input for class i is defined as

$$C_i \; = \; P_{SVM}^{Class}(i) = \frac{n_i}{K}; \;\; where \; i \; \in [1,N]. \tag{9}$$

We have stored P_{SVM} = $\{P_{SVM}^{Class}(1), \; P_{SVM}^{Class}(2), \; ...., \; P_{SVM}^{Class}(N)\}$ for the fusion. In the second phase of our algorithm, we have used unsupervised learning algorithm where no training is required on dataset D. In unsupervised approach, when unknown input of dimension $K \times P$ arrives to the system, it automatically employs Fuzzy C-Means clustering (FCM) algorithm to group all the observation (Dataset D plus $K \times P$ dimensional unknown input) into C (where $2 \leq C \leq (D_{dim} - 1)$) clusters. Basically, FCM provides us C cluster centers and a set of membership values. Then we have calculated Euclidean distance of each of K observations of the unknown input from these cluster centers. According to the Euclidean distance, we label them as cluster i $\forall \; i \in [1,C]$ e.g if the observation $O_r \; \forall r \in [1,K]$ is nearest to the cluster 3 then it is labeled as 3. Once the labeling is done for each of K observations, we generate a probability score based on the distribution of labeling for the unknown input among all the C clusters. If n_i $(n_i <= K)$ be the number of observations detected as cluster i, then the probability score of unknown input for cluster i is defined as

$$P_{FCM}^{Cluster}(i) = \frac{n_i}{K}; \;\; where \; i \; \in [1,C]. \tag{10}$$

Then in similar fashion, we store $P_{FCM} = \{P_{FCM}^{Cluster}(1), \; P_{FCM}^{Cluster}(2), \; ..., \; P_{FCM}^{Cluster}(C)\}$ for fusion process.

To apply Dempster-Shafer Theory (DST) for combining stored probabilistic mass P_{FCM} and P_{SVM}, we have defined frame of discernment (FOD) as

$$\begin{aligned} FOD = \theta_{person} = 1, 2, 3, ..N \;, \\ where \; N = total \; number \; of \; persons \end{aligned} \tag{11}$$

We have also done basic probability assignment P_{noise} (BPA) to θ_{person} based on noise present in the skeleton data. In our case, the P_{noise} is the degree to which the system fails to properly identify a person, but it has the knowledge that the person belongs to the FOD. It mainly occurs due to the noisy dataset. In our experiment, we have realized P_{noise} by considering height noise (outliers) of the dataset. Here, we assume that height of all persons should lie between 4feet 5inch to 6feet 3inch. So, any observation with a height beyond this range is treated as height-noise. We computed P_{noise} as the ratio of number of observation with height-noise with respect to the total number of observation. Moreover we consider that P_{noise} only occurs in the clustering process because SVM used to label each of K observations of the unknown input in anyone of the N classes so there is no chance of uncertainty in labeling. In formal notation, P_{noise} is defined as

$$P_{noise} = \frac{n_o}{(K + D_{dim})}; \tag{12}$$

where, n_o = total number of observation with height-noise.

In the last phase of our algorithm, we have applied DST to fuse P_{FCM}, P_{SVM} and P_{noise} to generate joint mass for i^{th} person using equation (13)

$$\begin{aligned} P_{fusion}(i) = K^{-1} \sum_{(P_{FCM} \cup P_{noise}) \cap P_{SVM} = i} (P_{FCM} + P_{noise}) * P_{SVM} \;\; \forall i \in [1,N] \\ where, \; K = 1 - \sum_{(P_{FCM} \cup P_{noise}) \cap P_{SVM} = \varphi} (P_{FCM} + P_{noise}) * P_{SVM} \end{aligned} \tag{13}$$

Entire flow of the algorithm is shown in the Figure 2 and Figure 3.

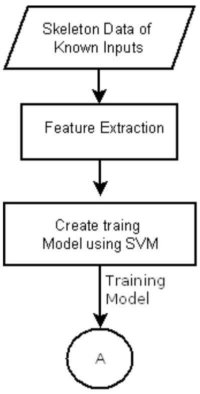

Figure 2: Flowchart of the proposed algorithm.

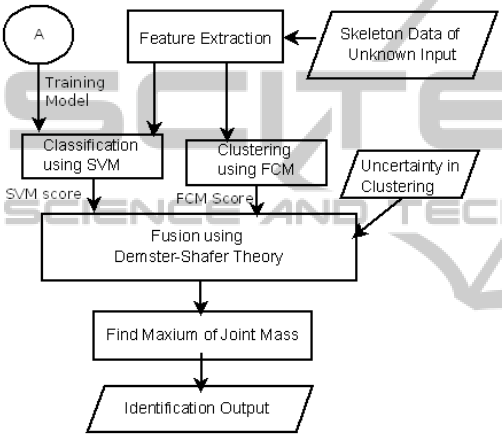

Figure 3: Flowchart of the proposed algorithm.

4 EXPERIMENTAL RESULTS

We have performed the experiments using 8 subject dataset (with 7 male and 1 female subjects) and considering two scenarios i) using singleton supervised or unsupervised algorithm based identification and ii) our proposed fusion (supervised + unsupervised) based approach. Throughout our experiments performance evaluation and comparison are done using precision and recall based F_{score} metric. All the experiments are performed in a Matlab environment using a Core-2 Duo Intel platform running at 2.53 GHz. We have used Windows SDK for Kinect version 1.6 (Microsoft, 2013) for skeleton data extraction and LibSVM (Chang and Lin, 2011) package for the classification.

Initially a kinect device has been used to record the skeleton data from a distance of 6 feet. Then feature extraction on each half-gait cycle of the 8 subjects (A-H) is performed resulting in a 46 dimensional feature vector. We store these feature vector for 8 persons as a dataset D. With this feature vector we perform the identification job. As mentioned in the Section 3.3, we first employ Support Vector Machine (SVM) classifier to generate a model using D as input. Now a new unknown test data is considered which needs to be identified from among the known 8 subjects. Feature vector for the test data is evaluated and SVM test is performed to obtain a set of probabilities. For e.g. when a new test data of dimension $O \times 46$ (where number of observation O = 39) for the subject H is taken, SVM testing provides us a set of probabilities (using equation (9)) mentioned in the Table 1.

Table 1: Confidence score using SVM classifier.

Class Label	SVM score (C_i)	Decision
A	0.028	
B	0.153	
C	0.111	
D	0.097	Detected
E	0.069	
F	0.084	
G	0.0	
H	0.458	

Now we have stored SVM score C_i $\forall i \in [1,8]$ in an array P_{SVM} for the fusion process. P_{SVM}=[0.028, 0.153, 0.111, 0.097, 0.069, 0.084, 0, 0.458].
In the next step, fuzzy c-means (FCM) clustering is performed with 3 clusters on dataset D along with the test data. Then labeling is done for each of the observation of test data with respect to its Euclidean distance from each of the centers. For our example, we have labeled each of O observations as cluster 1 or cluster 2 or cluster 3, according to its Euclidean distance from 3 cluster centers. In the other words, if any observation is nearer to cluster 1, it is labeled as 1. Once the labeling is done, we calculate probability of observation belonging to a particular cluster with respect to total number of observation. For our example case, the probabilities obtained for three clusters (using equation (10)), are tabulated in the Table 2. In similar fashion, FCM score are also stored in an array P_{FCM} to perform the work of fusion

Table 2: Confidence score using FCM.

Cluster Label	FCM Score ($P_{FCM}^{Cluster}(i)$)
Cluster 1	0.3073
Cluster 2	0.3072
Cluster 3	0.3473

P_{FCM}=[0.3073 0.3072 0.3473].
For the fusion process by Dempster-Shafer theory, we have also computed uncertainty by equation (12).
Uncertainty = P_{noise} = 0.0382.
Now, fusion score (joint mass) is computed by combining individual scores P_{SVM}, P_{FCM} and P_{noise} using equation (13). Fusion score for subject H is shown in Table 3.

Table 3: Fusion score for subject H.

Class Label	Fusion score ($P_{fusion}(i)$)	Decision
A	0.017	
B	0.072	
C	0.072	
D	0.098	Detected
E	0.022	
F	0.02	
G	0.024	
H	0.675	

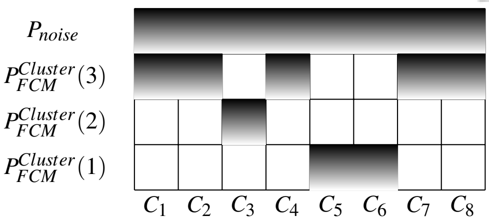

Figure 4: Fusion using Dempster-Shafer Theory.

K mentioned in the equation (13) is the sum of area of the shaded portion shown in the Figure 4 and calculated as K = 0.3699.

It can be observed from Table 3, that the unknown input pattern is not only correctly identified as Person 8 (subject H), with fusion score $P_{fusion}(8) = 0.675$, but also has better confidence score than singleton SVM classifier(mentioned in Table 1 as C_8 = 0.458). So, experimental results proves that our proposed fusion algorithm using Dempster-Shafer theory is performing better and able to identify the person H with improved level of accuracy.
Next we analyze the overall improvement in confidence score of recognition using the confusion matrices for the above mentioned two approaches namely with fusion and without fusion i.e. only using SVM as classifier.

The resulting recognition accuracy using SVM in terms of F_{score} is shown in the Table 4. Then we have applied our proposed algorithm under the same circumstances and the results are presented in the Table 5.

Table 4: Confidence score using SVM for different test subjects (green means detected, TS= Test Subject and GT = Ground Truth).

TS \ GT	A	B	C	D	E	F	G	H
A	0.590	0.026	0.0	0.076	0.0	0.103	0.205	0.0
B	0.204	0.510	0.082	0.102	0.0	0.020	0.0	0.082
C	0.116	0.256	0.419	0.070	0.070	0.0	0.023	0.046
D	0.071	0.095	0.095	0.667	0.024	0.048	0.0	0.0
E	0.0	0.121	0.030	0.030	0.769	0.0	0.0	0.061
F	0.054	0.0	0.0	0.162	0.0	0.730	0.054	0.0
G	0.037	0.0	0.0	0.0	0.0	0.0	0.963	0.0
H	0.028	0.153	0.111	0.097	0.069	0.084	0.0	0.458

Table 5: Fusion score for different test subjects (green means detected, TS= Test Subject and GT = Ground Truth).

TS \ GT	A	B	C	D	E	F	G	H
A	0.617	0.023	0.0	0.075	0.0	0.100	0.185	0.0
B	0.144	0.573	0.092	0.083	0.0	0.016	0.0	0.092
C	0.119	0.293	0.427	0.041	0.041	0.0	0.026	0.053
D	0.038	0.071	0.050	0.759	0.028	0.054	0.0	0.0
E	0.0	0.111	0.028	0.031	0.769	0.0	0.0	0.061
F	0.009	0.0	0.0	0.116	0.0	0.836	0.039	0.0
G	0.009	0.0	0.0	0.0	0.0	0.0	0.991	0.0
H	0.017	0.072	0.072	0.098	0.022	0.02	0.024	0.675

Diagonal of these two tables (Table 4 and Table 5) clearly helps us to understand that the fusion of SVM and FCM score improves the recognition accuracy for all the subjects. The recognition accuracy with fusion and without fusion for all the subjects are summarized in terms of average F_{score} in Table 6. Therefore one can clearly infer that the proposed fusion based method outperforms any existing singleton approach for person identification. One thing that needs to be mentioned here is that we have not reported the performance metric of person identification using only unsupervised algorithm because the reported recognition accuracy by Ball et.al in (Ball et al., 2012) for only four persons using K-Means algorithm is only 43%.

Table 6: Performance comparison between without fusion and with fusion.

Without Fusion	With Fusion
0.64	0.71

5 CONCLUSIONS

In this paper we have proposed a novel approach based on fusion of supervised and unsupervised learning algorithm using Dempster-Shafer theory in defining the final decision metric of human identification. Results indicate that the combination of our proposed fusion algorithm outperforms existing framework of person identification in real time. As a future work, we would be experimenting on gait independent features, which would further improve the robustness of the system by getting rid of gait boundary detection, as well as remove the constraint on the side walk.

REFERENCES

Ball, A., Rye, D., Ramos, F., and Velonaki, M. (2012). Unsupervised clustering of people from 'skeleton' data. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction (HRI)*, pages 225–226.

Begg, R. K., Palaniswami, M., and Owen, B. (2005). Support vector machines for automated gait classification. *IEEE Transactions on Biomedical Engineering*, 52(5):828–838.

BenAbdelkader, C., Cutler, R., and Davis, L. (2002). Stride and cadence as a biometric in automatic person identification and verification. In *Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 372–377.

Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algoritms*. Plenum Press, New York.

Bezdek, J. C., Ehrlich, R., and Full, W. (1984). Fcm: The fuzzy cmeans clustering algorithm. *Computers & Geosciences*, 10(2):191–203.

Bouchrika, I. and Nixon, M. S. (2008). Exploratory factor analysis of gait recognition. In *8th IEEE International Conference on Automatic Face & Gesture Recognition, 2008. FG'08*, pages 1–6. IEEE.

Carlsson, S. (2000). Recognizing walking people. In *Proceedings of the 6th European Conference on Computer Vision (ECCV) -Part I*, pages 472–486.

Chang, C.-C. and Lin, C.-J. (2011). Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27.

Cheng, L., Sun, Q., Su, H., Cong, Y., and Zhao, S. (2012). Design and implementation of human-robot interactive demonstration system based on kinect. In *Control and Decision Conference (CCDC), 2012 24th Chinese*, pages 971–975. IEEE.

Cheng, M.-H., Ho, M.-F., and Huang, C.-L. (2008). Gait analysis for human identification through manifold learning and hmm. *Pattern recognition*, 41(8):2541–2553.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.*, 20(3):273–297.

Dadashi, F., Araabi, B., and Soltanian-Zadeh, H. (2009). Gait recognition using wavelet packet silhouette representation and transductive support vector machines. In *2nd International Congress on Image and Signal Processing, 2009. CISP '09*, pages 1–5.

Dempster, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *The annals of mathematical statistics 38*, (2):325–339.

Dempster, A. P. (1968). A generalization of bayesian inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 205–247.

Fine, T. L. (1977). Review: Glenn shafer, a mathematical theory of evidence. *Bulletin (New Series) of the American Mathematical Society 83*, (4):667–672.

Huang, P., Harris, C., and Nixon, M. (1999). Human gait recognition in canonical space using temporal templates. *IEE Proceedings- Vision, Image and Signal Processing*, 146(2):93–100.

Jeffreys, H. (1973). *Scientific Inference (3rd ed.)*. Cambridge University Press p. 31.

Jsang, A. and Pope, S. (2012). Dempsters rule as seen by little colored balls. *Computational Intelligence 28*, (4):453–474.

Kale, A., Cuntoor, N., Yegnanarayana, B., Rajagopalan, A., and Chellappa, R. (2003). Gait analysis for human identification. In *Audio-and Video-Based Biometric Person Authentication*, pages 706–714. Springer.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME Journal of Basic Engineering*, (82 (Series D)):35–45.

Karem, F., Dhibi, M., and Martin, A. (2012). Combination of supervised and unsupervised classification using the theory of belief functions. In *Belief Functions: Theory and Applications*, pages 85–92. Springer.

Le Hegarat-Mascle, S., Bloch, I., and Vidal-madjar (1997). Application of dempster-shafer evidence theory to unsupervised classification in multisource remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 35(4):1018–1031.

Meyer, D., Psl, J., and Niemann, H. (1998). Gait classification with hmms for trajectories of body parts extracted by mixture densities. In *British Machine Vision Conference*, pages 459–468.

Microsoft (2013). Kinect for windows. http://www.microsoft.com/en-us/kinectforwindows/develop/developer-downloads.aspx. [Online; accessed 25-July-2013].

Pal, N. R. and Chintalapudi, K. K. (1997). A connectionist system for feature selection. *Neural Parallel Sci. Comput*, 5(3):359381.

Preis, J., Kessel, M., Werner, M., and Linnhoff-Popien, C. (2012). Gait recognition with kinect. In *1st International Workshop on Kinect in Pervasive Computing*.

Sarkar, S., Phillips, P., Liu, Z., Vega, I., Grother, P., and Ortiz, E. (2005). The humanid gait challenge problem: data sets, performance, and analysis. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 27, pages 162–177.

Shafer, G. (1976). *A mathematical theory of evidence Vol. 1*. Princeton: Princeton university press.

Sinha, A., Chakravarty, K., and Bhowmick, B. (2013). Person identification using skeleton information from kinect. In *The Sixth International Conference on Advances in Computer-Human Interactions (ACHI)*, pages 101–108.

Wang, J., She, M., Nahavandi, S., and Kouzani, A. (2010). A review of vision-based gait recognition methods for human identification. In *International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, pages 320 – 327.

Xue, Z., Ming, D., Song, W., Wan, B., and Jin, S. (2010). Infrared gait recognition based on wavelet transform and support vector machine. *Pattern Recogn.*, 43(8):2904–2910.