ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

A Unified Framework for Human Activity Detection and Recognition for Video Surveillance Using Dezert Smarandache Theory.

Srilatha V, and Veeramuthu Venkatesh*

School of Computing, SASTRA University, Tamil Nadu, India.

ABSTRACT

Trustworthy contextual data of human action recognition of remotely monitored person who requires medical care should be generated to avoid hazardous situation and also to provide ubiquitous services in home-based care. It is difficult for numerous reasons. At first level, the data obtained from heterogeneous source have different level of uncertainty. Second level generated information can be corrupted due to simultaneous operations. In this paper human action recognition can be done based on two different modality consisting of fully featured camera and wearable sensor. Computationally event features are got from the images given by camera and movement actions are provided by wearable sensor. Human action realization, we have to use both decision and feature level fusion methods are studied by a collaborative classifier. By using feature level method inputs from different sources are combined before going to classification action. For decision level fusion DsMT is used to combine the outputs from two classifiers, each corresponds any one of the sensor. The proposed frame works is validated using Berkeley Human action database. Based on this frame work human action recognition can be done effectively with increased level.

Keywords: Wireless sensor networks, Activity Recognition, Senor Fusion, Dempster Shafer theory, Dezert Smarandache theory, video surveillance.

Corresponding author

INTRODUCTION

Human action recognition [1-2] application can used along with human computer interaction (HCI), it also include for application like surveillance, elderly people monitoring, and context aware computing. HCI is also used for rehabilitation and body fitness training. While considering the human action recognition makes use of camera and wearable senor.

Microsoft introduce an in depth camera called Kinect will help us to monitor the action done by any people inside and also outside. With the help of light depth sensor depth image can generated and structured whereas Kinect camera are very insensitive to changes in light conditions and also it is difficult to provide 3D information towards distinguishing action. Normally an action graph was introduced to model the dynamic action and a 3-D point from in depth images was depleted to characterize postures. Body shape is represented by depth motion map (DMM) method and human action and movement information is followed by support vector machine principle. Random occupancy features and patterns are extracted from weighted sampling method. Numerous action recognition models includes wearable sensor [3-4]. These sensors are used to recognize day to day activities by using a method called artificial neural network (ANN) within the tree based structure. Human daily activity and recognition was introduced by sparse representation classifier. Normally wearable inertial sensor and depth sensor are used to monitor the human action recognition. In this paper, we considering both decision level method and feature level method are considered. The decision level fusion [5-6] method is performed by DsMT [7-8] method. The new fusion approach is valued by using a publically available data base called multi modal human action database and Berkeley model. The performance is can be compared in any situations by using modality sensor individually. Most of the time in darkness depth and wearable sensor is used due to less cost and also easy to operate.

Mathematical Techniques:

Sparse Representation Classifier (SRC):

The information provided for classifying the images starting from its least mean square formations are usually represented by the set of dictionaries or classes which was denoted as'd'. The decision making through sparse representation based classifier (SRC) is represented as S_d

$$S_d = \| x - Y_d \hat{\alpha}_d \| \qquad \text{------ (1)}$$

Where $\hat{\alpha}_d$ is the coefficients and Y_d are the inputs samples for dictionary d represented as x which is defined by the coefficients of $\hat{\alpha}$ and the training samples as Y. The decision making can be done in two class's minimum class and maximum class the minimum class is represented as

$$\text{Class}(n) = \arg_d \min S_d(n) \qquad \text{------ (2)}$$

$$\text{Class}(n) = \arg_d \min \| \hat{\alpha}_d \| \qquad \text{------ (3)}$$

In SRC classification, the dictionary samples are represented by maximal and minimal coding residuals of an image. It does not represent the combination of query samples that are ignored. The use of all dictionary samples are collaboratively represented by Collaborative Representation Classifier so that the decision making will be more accurate and also can avoid uncertainty.

Collaborative Representation Classifier (CRC):

The special case of CRC is applying different characterization of coding which is related to robustness of CRC to the pixels of obtained input image and the verification of activity will be more precise and robust the regularized least squares can be obtained by CRC are represented by

$$\hat{\alpha}_{CR} = P_y;\ P = (Y^T Y + \hat{\lambda}_{CR} 1)^{-1} Y^T \qquad \text{------ (4)}$$

$$R_d = \frac{x - Yd\ \hat{\alpha}\ d\|}{\hat{\alpha}\ d\|} \qquad \text{------ (5)}$$

Where R_d is the combination of all query samples and the dictionaries which are ignored in sparse representation classifier and these are calculated by the collaborative representation classifier. The decision making is more precise when the input images are regularized using CRC than SRC technique.

Dempster Shafer Theory (DST) [9-10]:

DST was introduced by Dempster and further developed by Shafer the main goal is to deal with uncertainty and imprecision of input data provided by sensors and can be effectively used for data fusion applications. Let Θ be a finite universal set of both union and intersection input values, which is called as an individual frame of mass functions. In fusion applications, Θ corresponds to a set of classes. The exponential finite function 2^{Θ} is the set of all possible input frames of Θ. A mass input function is defined by the probability of $m: 2^{\Theta} \rightarrow [0, 1]$.

$$\sum_{S \subseteq \Theta} m(S) = 1 \quad m(\Phi) = 0 \qquad \text{------ (6)}$$

Where Θ is the empty set, and S is the subset with non zero input mass elements. The value of m (S) is a measure of the belief that is assigned to set S, not to subsets of S. Two common elements which measures the belief, and plausibility functions, are can be represented as (S $\subseteq$ Θ, $R \subseteq$ Θ)

$$\text{Bel}(S) = \sum_{R \subseteq S} m(R) \qquad \text{------ (7)}$$

$$\text{Pl}(S) = \sum_{R \cap S \neq \phi} m(R), \ \text{Pl}(\phi) = 0 \qquad \text{------ (8)}$$

$$\text{Bel}(A) \leq \text{Pl}(A), \ \text{Pl}(A) = 1\text{-Bel}(A') \quad \text{------ (9)}$$

The two values or mass functions can be obtained by using dempster's rule

$$K = \sum_{B \cap C = \phi} m1(B)m2(C) \qquad \text{------ (10)}$$

Where K is the normalization factor which provides a combination of belief and plausibility factors measured for two input values or mass functions. This rule is commutative and associative. If there are more than two input sources or mass functions, the combination rule can be generalized by iteration.

Dempster Shafer theory limitation

The statement which contradicts itself and yet might be true or wrong at the same time with the different degrees of uncertainty of a data is avoided by current method using DST and combining those evidences obtained by the DST is further developed by DsMT.

Dezert Smarandache theory (DsMT):

The Dezert Smarandache theory of likely and paradoxical reasoning of a simple continuation of basic Dempster shafer theory but the difference which includes the combination of individual input mass elements in a particular information which is represented as a belief functions, but it mainly focuses on fusion of uncertainty and highly paradoxical characteristics and imprecise of both the quantitative and qualitative input elements specially it deals with the vague input mass functions.

Let θ = {θ 1 . . . θ n} be a set of n input mass functions which is considered as union that cannot be defined precisely and let's consider the basic elements $\bigcup$ (union) and $\bigcap$ (intersection). The θ will be called a basic input mass function. But this new theory can be added as the Dempster Shafer theory even further by accepting the possibility for paradoxical information.

$$m(\theta 1) + m(\theta 2) + m(\theta 1 \cup \theta 2) + m(\theta 1 \cap \theta 2) = 1 \quad \text{------ (11)}$$

Patient Action Database:

The Berkeley Human action database is used to synchronize all the captured images or frames from a video which is recorded from the kinect depth cameras, mainly for the purpose of storing the information gathered. The collected information is incorporated by the variations of activities performed by the person. The recorded actions like Sleep (Sl), Fall (F), Sit (S). The collected information is organised so that it can be easily accessible, manageable, and can be modified or updated accordingly by the user. This Berkeley action database contains the aggregation of data records which manages the relevant information needed by the user.

Implementation:

Activity recognition of a person in a home based care environment frame work is shown in figure 1 using both wearable sensors and camera in this paper we are considering wearable sensors as Accelerometer sensor for identifying the physical activates, fibre optical sensors for posture recognition, pressure sensor used to identify the respiration or pulse rate of a person and body temperature sensor is used to calculate the body temperature. Here the use of camera is to identify the movement actions. The obtained sensor events are further extracted using the combination of both the Dempster shafer theory and Dezert Smarandache theory. The camera inputs are taken from the Berkeley Human action Data base the inputs stored in data base are numerical values of images which are captured from the camera. The fusion technique can be done by using Sparse Representation Classifier and Collaborative Representation Classifier techniques these can be extracted using Dempster Shafer Theory. The obtained outputs are identified and the decision making of a particular activity is recognised.

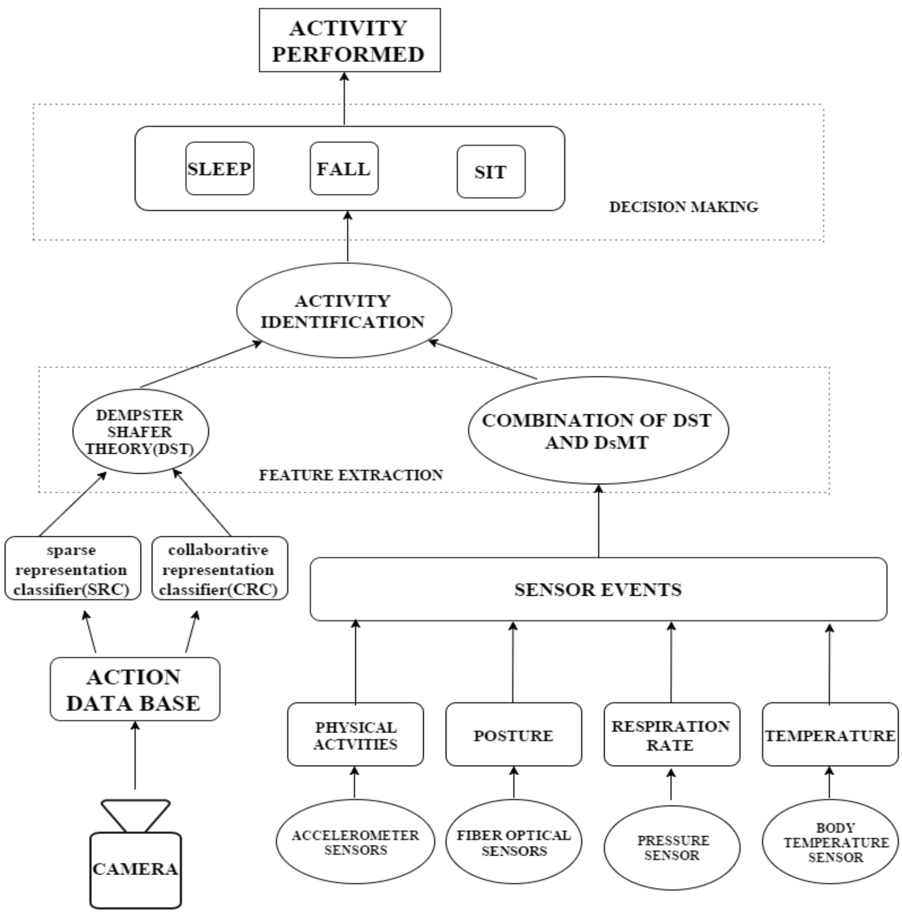

Figure: 1 Frame work for Activity Recognition

Fusion Extraction: (SRC & CRC)

Table 1 represents the input values of stored images of user activity which is taken from Berkeley human action data base

Mass functions	Sleep	Fall	Siting position
Dictionary inputs (x)	1.7	0.66	1.67
Training samples ($Y_d \hat{\alpha}_d$)	1.2	3	1

Table 1: Mass function values obtained by camera which is stored in Berkeley Human action database

In feature level of fusion the input values are first applied to SRC and CRC to the depth images obtained, these obtained values are considered as mass functions for DST for further decision making of identifying the activity recognition of a particular patient. Applying SRC to the above input values, the obtained values after SRC extraction is represented in table 2

Sparse Representation Classifier	Sleep	Fall	Siting position
S_d	0.5	0.33	0.67

Table 2: Obtained values after applying Sparse Representation Classifier

Applying CRC to the above input values which is given in table 1 and after the extraction of CRC the values are represented in table 3

Collaborative Representation Classifier	Sleep	Fall	Siting position
R_d	0.75	0.11	0.67

Table: 3 Obtained values after applying Collaborative Representation Classifier

After the fusion extraction the final result which is extracted from the inputs of camera is again fused using Dempster shafer theory and the obtained values are represented as table 4, which represents the obtained mass functions of Dempster shafer theory.

Dempster Shafer Theory	Sleep	Fall	Siting position
M	0.3404	0.15	0.693

Table: 4 Obtained values after applying DST

The decision making after applying SRC and CRC result shows as the patient action is recognised as sitting clearly.

Fusion Extraction: (DST & DsMT)

Mass functions	Sleep	Fall	Siting position
M1	0.0001	0.0005	0.001
M2	0.01	0.05	0.1
M3	0.5	0.001	0.0001
M4	0.0001	0.005	0.1

Table: 5 mass function values obtained by sensors

Table 5 represents the input mass function values obtained by wearable sensors. Applying DST to the above input values

M (Sl) = 0.000005, M (F) = 0.000005, M (S) = 0.000001 ------ (12)

Examine a patient and agree that it monitors from Sleep (Sl), or Fall (F) Thus θ = {Sl, F}. Assume that the doctors agree in their low expectation of a Sit, but disagree in likely cause.
Applying the DsMT to the fused values of DST
Bel (Sl) = m (Sl ∩ F) + m (Sl ∩ S) = 0.00005
Bel (F) = m (Sl ∩ F) + m (S ∩ F) = 0.00005
Bel (S) = m (S) + m (Sl ∩ F) = 0.00001 ------ (13)

If the doctors can be resulted as same condition, then the combined information mass elements are mainly focused on the weight of occurrence and on the paradoxical situation as both sleep and fall Sl ∩ F which represents that patient is in the conditions as sleeping and fallen but not in the condition as sitting. This is clearly stated as the paradoxical situation. In this case, the DST had concluded that the patient is in the condition of either sleeping or fallen with certainty. Applying DsMT to the above input values we will have no change in frame of discernment which was the collection of documents. Then each individual inference kind corresponded to every query term, will be a body of evidence and its mass function will be amount of evidence raised by that inference on a document.

M(S) = $\sum_{S \subseteq Sh} m\ (S)\ *\ \mathrm{m}\ (\mathrm{Sl})$------ (14)
M (S) > 0.000005 ------ (15)

Comparing all the values with the remaining obtained DST fused values the mass function of the position sitting has more probability of occurrence. We can be no longer worry about paradoxical situation because we now have an interpretation for that. So the decision making after applying DsMT result shows as the patient action is recognised as sitting clearly. By comparing both the sensory and camera output shows the result as the patient is doing the activity as sitting clearly this method is more efficient for activity recognition as 80% and can be mainly used for elderly people in hospitals or in home environment for multiple persons without any caretakers.

Conclusion and Future Enhancements.

In this paper we have presented an ontological framework to aggregate sensor and video data to resolve of activity recognition in a smart home environment. The proposed novel approach uses an ontology to define how the sensor and video data are correlated to the activity, the person and the objects within the environment. The sensor and video data were synchronized to determine how the sensor and video measures were related. The results from initial experiments show that video actions provided additional information relating to the location. In addition, the video events can be used to improve the sensor events and deliver a greater accepting of the activity being performed. Furthermore, the video actions can be used to overwhelmed problems associated with differences in the sensor data such as missing data. The proposed method has concentrated on a single entity being performed by a single user, though, this solution is inadequate and the work will be prolonged in the future to incorporate a determination for a single person accomplishment multiple activities, several people performing a single activity and several people performing numerous activities.

REFERENCES

[1] Ahouandjinou, A. S. R. M., Motamed, C., & Ezin, E. C. (2015). A temporal belief-based hidden markov model for human action recognition in medical videos. *Pattern Recognition and Image Analysis, 25*(3), 389-401.

[2] Jiang, Y. -., Dai, Q., Liu, W., Xue, X., & Ngo, C. -. (2015). Human action recognition in unconstrained videos by explicit motion modeling. *IEEE Transactions on Image Processing, 24*(11), 3781-3795.

[3] Kunze, K., & Lukowicz, P. (2014). Sensor placement variations in wearable activity recognition. *IEEE Pervasive Computing, 13*(4), 32-41.

[4] Lo, G., González-Valenzuela, S., & Leung, V. C. M. (2013). Wireless body area network node localization using small-scale spatial information. *IEEE Journal of Biomedical and Health Informatics, 17*(3), 715-726.

[5] Belde, U., & Lebleb Iciog lu, M. K. (2015). A new systematic and exible method for developing hierarchical decision-making models. *Turkish Journal of Electrical Engineering and Computer Sciences, 23*(1), 279-297.

[6] Sannen, D., & Van Brussel, H. (2012). A multilevel information fusion approach for visual quality inspection. *Information Fusion, 13*(1), 48-59.

[7] Jamrozik, W. (2015). Contextual reliability discounting in welding process diagnostic based on DSmT. *Expert Systems, 32*(2), 192-202.

[8] Khodabandeh, M., & Shahri, A. M. (2014). Uncertainty evaluation for a dezert-smarandache theory-based localization problem. *International Journal of General Systems, 43*(6), 610-632.

[9] Chen, C., Jafari, R., & Kehtarnavaz, N. (2015). Improving human action recognition using fusion of depth camera and inertial sensors. *IEEE Transactions on Human-Machine Systems, 45*(1), 51-61.

[10] Zhang, F., & Ge, Z. (2015). Decision fusion systems for fault detection and identification in industrial processes. *Journal of Process Control, 31*, 45-54.