The ICCIT International Conference
Proceeding

IEEE Conference Record number: #17832
IEEE PDF Files Catalog Number: CFP1032I-ART
IEEE PDF Files ISBN: 978-89-88678-30-5
IEEE PDF Files Catalog Number: CFP1032I-PRT
IEEE PDF FilesISBN: 978-89-88678-29-9

ICCIT 2010 Vol. I

The 5th International Conference on Computer Sciences and Convergence Information Technology

November 30 - December 2, 2010,
Grand Hilton Hotel, Seoul, Korea

Sponsored by

IEEE · IEEE SEOUL SECTION · AICIT

Hosted/Co-organized by

ETRI · IEEE IRAQ SECTION

5th International Conference on Computer Sciences and Convergence Information Technology

ICCIT2010

Table of Contents

A Multi-Strategy Bayesian Model for Sensor Fusion in Smart Environments

Muhammad Fahim
Ubiquitous Computing Lab, Kyung Hee University, Korea
fahim@oslab.khu.ac.kr

Muhammad Hameed Siddiqi
Ubiquitous Computing Lab, Kyung Hee University, Korea
siddiqi@oslab.khu.ac.kr

Sungyoung Lee
Ubiquitous Computing Lab, Kyung Hee University, Korea
sylee@oslab.khu.ac.kr

Young-Koo Lee
Ubiquitous Computing Lab, Kyung Hee University, Korea
yklee@khu.ac.kr

***Abstract*—Sensor fusion became a powerful scheme to recognize the daily life activities in smart homes. This paper proposed a multi-strategy approach to overcome the challenges of accuracy and efficiency. We design a model to integrate k-Nearest Neighbor (kNN, k=5) technique and Bayes classifier for recognizing the activities of daily living. There are three stages of this model. The first stage is used to reduce the search space by discovering the useful regions. A Bayes classifier is utilized in the second stage to refine the degree of beliefs. The confidence values have been denoted by the output of the Bayes classifier. Finally, max rule has been applied to fuse confidence values. The proposed model has been evaluated on five different types of activities from Place Lab dataset (PLIA1). We compare our Multi-strategy approach with the Naïve Bayes Classifier and get 9% higher accuracy and 186 ms faster execution time.**

Keywords-component; k-Nearest Neighbor (kNN); Bayesian Classifier; Sensor Fusion.

I. INTRODUCTION

Smart environments focus on supporting the daily life activities. The nomenclature of Smart environments has two major spaces, Ambient Spaces and Smart Spaces [1]. Depending on the purpose, ambient spaces have different computing configurations. In Ambient Spaces, sensors are placed at exact locations to support the activities of daily living. Locating the sensors becomes a big challenge for the "whole-home" environment [2]. Smart Spaces are well equipped with the heterogeneous types of sensors. They also support the ability to transfer the computational resources at any point within the environment [1]. Furthermore, they support the true setup without disturbing the user's daily life and are totally invisible. Recognizing which activity is performing is a complex and challenging task in the Smart Spaces.

Sensor fusion plays an important role in recognizing the daily activities to fuse the different sources of information. The collected information from individual sensors has different representation, different level of abstraction and diverse in nature [3]. In the past, the researchers have investigated in fusion techniques; e.g., agent based communities [4], Bayesian fusion by local approaches [5, 6] and rough set method [7]. We have also proposed Multi-strategy Bayesian approach for sensor fusion. In this model, heterogeneous sensors data is fused to recognize the activities.

We used the dataset in which environment is continuously monitored. A bulk of information is collected after some instance of time. To process the cumbersome information has high computational cost. By applying the kNN technique, relevant features have been highlighted as a useful region. It reduced the computational cost. It made the basis for Bayesian inference process and corresponds to the base classifier.

The Bayesian approach is a powerful probabilistic model for fusing the homogeneous and heterogeneous information [5]. Estimated probabilities depend upon the information sense by the different sensors. The complexity of Bayesian model tasks increases exponentially with the number of sources [8]. For getting the high accuracy to recognize the activities it requires the initial knowledge of many probabilities. In this process degree of belief about the certain type of hypothesis is built. Finally, the confident results of each sensor are selected by the max rule.

Our focus is to explore the Smart Spaces with respect to supporting the activities of daily living. Our approach is able to recognize the activities accurately and provides the results in reasonable time. The objective of this paper is to resolve the two major challenges including the accuracy and computational cost.

For empirical evaluation, we implemented our idea on the standard place lab dataset. Our results showed that the multi-strategy approach is more accurate and efficient. We believe that our experimental test bed is useful for further research in the area of sensor fusion. In the context of this paper we proposed our initial model for sensor fusion.

We structured our paper as followings: Section II presents some of the existing approaches for Bayesian inference modeling. Section III presents our new approach for sensor fusion. In Section IV, we introduce our practical results followed by discussion. And finally the conclusion and future work are drawn in Section V.

II. RELATED WORK

The authors of [5] proposed a framework to overcome the challenge of heterogeneity of sensors information representation. They provided a technique to handle with the uncertain, redundant and time sensitive information. In their technique, agents communicated with each other and posse's useful information for decision making purposes.

The authors of [9] presented a local approach for Bayesian fusion. Local Bayesian fusion based on coarsening and restriction techniques. They focused on coarsening and modeled successfully their architecture on the real world criminal investigation process. Their technique was valid in the both top down and bottom up approach for building the fusion model. The drawback of this technique is finding out the quantities of interest locally so global view space which contains the useful information may be missed.

The authors of [5] also presented a local Bayesian fusion approach for the reduction of storage and computational cost. Task specific prior knowledge and sensory information were used to build local Bayesian setups in a flexible and problem specific manner. Using the concept of misleading evidence, the most suspicious elements of the space were identified. They claimed that their technique is more flexible by creating the setup for fusion under the consideration of all information.

The authors of [10] rephrased this problem in terms of the minimization of an objective function. They developed IN and DN methods to improve the results. In the IN method they assumed that the validation variables are conditionally independent. While in DN method these variable are dependent on each other.

The authors of [7] used rough set theory for attribute reduction of the data. First the uncertain and redundant information of the dataset were removed. Attributes were reduced by calculating the importance of each attribute. Finally, they applied classifier fusion method for high accuracy in the results.

Some of the existing work [5, 9] have the problem of accuracy because they search information locally. The purpose of this paper is to develop an efficient model for recognizing daily life activities in Smart Environments. This model was evaluated on five different types of activities and produced reliable accuracy and efficiency.

III. THE PROPOSED APPROACH

The proposed approach for recognizing the activities of daily living is comprised of three stages. At the first stage we applied partially k-Nearest Neighbor technique. The objective of this stage is to reduce the number of classes and to get those feature vectors in search space which are more relevant with the current feature vector. At the second level prior and posterior probabilities of the useful regions are calculated to update the degree of beliefs. Finally, we use the max rule to fuse the information for decision making purpose. Fig. 1 shows the complete architecture of the proposed model.

A. Discovery of useful Regions

It is impossible to extract the desired information precisely and completely from a single source [5]. Information is scattered in the whole search space. It is important to find all those regions which contain the required information. It reflects the relationship between the current instance and training instances of sensory information. Instance-based learning techniques are commonly used such as nearest neighbor and locally weighted regression [11]. k-Nearest Neighbor technique has two stages; the first is the determination of the nearest neighbors and the second is the classification using those neighbors [12]. The uniqueness of our proposed approach is that, classification is the slow process in this technique. That

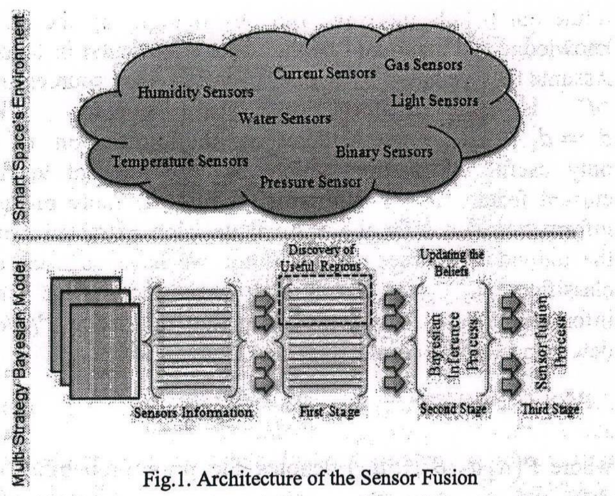

Fig.1. Architecture of the Sensor Fusion

is the reason why we have not performed classification at preprocessing stage. We only find the relevancy between the instances and take those one which are more similar with the current one. Assume "f_v" is the feature vector, which discovers the most relevant instances in complex feature space"f_s".

$$f_v(x_i) \leftarrow f_s(x_n) \quad (1)$$

Euclidian distance has been used to find the relevancy of real value instances [11]. By calculating the distance of "f_v" to instances of "f_s", it discovers those feature vectors which are closer to the relevant features.

$$d_{ij} := \sqrt{\sum_{k=1}^{n} (f_v(x_{ik}) - f_s(x_{jk}))^2} \quad (2)$$

Distance is calculated by the i[th] feature vector x_i with the training examples x_j. Based on this structure most relevant instances are obtained.

```
Input: training set T, current instance x, relevancy threshold
R, number of sensors S
Output: relevancy matrix
Begin
  while each Training Set T, do
    compute (index,S) := square(sensorData(index,S)-
    x(index,S))
    compute (index,S) := Σ_{i=1}^{n}(T, x_i)
  end
  compute := sort (compute,S+1)
  relevancyMatrix := Top R rows (compute)
End
```

Fig.2. Algorithm for useful regions

B. Bayesian Inference Process

After finding out the more relevant instances, useful regions are highlighted. For gaining the confidence, we refine our beliefs more precisely by multiplying the prior knowledge and maximal likelihood for a particular instance. Assume that we have *"S"* number of information sources and *"d"* is the information that contains like $d := d_1, d_2, d_3, \ldots, d_s$. At this stage, the information *"d"* is only useful information which is more relevant to the current feature vector. We want to find the more precise information and infer the optimal decision extracted from the individual sources *"S"*. Assume we have *"C"* sets of classifiers $(C_1, C_2, C_3, \ldots, C_n)$ and want to prune the information for these classes. Then the confidence *P(C/d)* determined by the certain observed vector d is:

$$P(C|d_1, d_2, d_3 \ldots d_n) = \frac{P(d|C_1, C_2, C_3, \ldots, C_n)P(C)}{P(d_1, d_2, d_3, \ldots d_n)}, \quad (3)$$

where $P(d_1, d_2, d_3, \ldots d_n)$ denotes the prior probability of data and we drop this, because it is a constant and independent of class labels. By this way, it brings about:

$$P(C|d_1, d_2, d_3 \ldots d_n) = P(d|C_1, C_2, C_3, \ldots, C_n)P(C). \quad (4)$$

Thereby, $P(d|C_1, C_2, C_3, \ldots, C_n)$ is the likelihood function. Confidence is the product of prior knowledge and maximal likelihood information of each sensor.

```
Input: Relevancy Matrix RM
Output: Confidence against each sensor with the class label
Begin
  for Each sensor in RM, do
    uAttribute = unique(RM(data,sensor))
    for Each uAttribute in sensor, do
      Calculate prior Probabilities
      Calculate maximal likelihood Probabilities
    end
    confidenceMatrix←prior*maximal likelihood
  end
End
```

Fig.3. Algorithm for Bayesian Inference

C. Final Fusion Step

The output of the Bayesian inference process is the confidence value and it has very clear inference about the class labels. Input sources are independent of each other and contain the confidence about the certain classes. Several fixed combining rules are used but they depend on output values of the base classifier [13]. As we have *"C"* classes (after getting the useful regions) and each class has *"T"* confidence value for different classes. We select the confidence of each class against "f_v" with the help of max rule as below:

$$\sum_{i=0}^{n} D_v = \max\left(T_i(C_j)\right). \quad (5)$$

In Eq.5 is the decision vector which selects the most confident value from each prune class.

IV. Results and Discussion

For recognizing the daily life activities in the Smart Spaces, we fused sensor information for this purpose. We use Place Lab Intensive Activity 1 (PLIA1) dataset that was recorded in MIT Place Laboratory [8]. Place lab is the real smart home where everyday activities can be observed and recorded for the experiments. It is true environment for developing the ubiquitous computing technology. Sensing devices are integrated into the architecture of the home. It contains one bedroom, dining area, kitchen, two bathrooms and small office [8]. A volunteer was performing the common house hold activities during the four hour period. In that time, volunteer preparing two recopies, dish washing, cleaning, laundering, office working, studying and light cleaning around the apartment [2]. For our experiment, we recognized five macro activities as depicted in Table I.

TABLE I. List of Activities

Daily life activities
Relaxing
Dish Washing
Meal Preparation
Laundry
Cleaning

We use the 7 different types of environment sensors data place at 44 different locations. Table 5 shows the name of the sensors and their quantities.

TABLE II. List of Environment Sensors

Type of Sensors	Qty
Humidity Sensors	10
Light Sensors	6
Temperature Sensors	6
Pressure Sensor	1
Current Sensors	8
Water Sensors	8
Gas Sensor	1

Humidity sensors measure in RH (Relative humidity) between 0% and 100%. There are 10 active humidity sensors located at different positions. Fig.4 shows the plotting of humidity sensor data. Descriptions about the dataset are briefly explained in the dataset documentation which is not the part of this paper discussion. Light Sensors measure irradiance in between 0.020 to 1.21 and its unit ismW/cm^2. Fig.5 shows the plotting of light sensor data. Total seven types of sensors were placed during the collection of dataset.

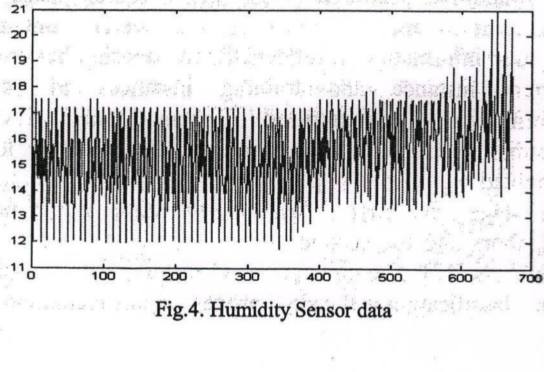

Fig.4. Humidity Sensor data

All the sensors have different measurement units and different scales.

In Fig. 4 and 5 show the data visualization of the one hour data that have been collected by the Humidity and Light sensors, respectively. Before we present our Multi-strategy approach recognition results, let us show the single classifier recognition for daily life activities recognized by the simple Naïve Bayes Classifier.

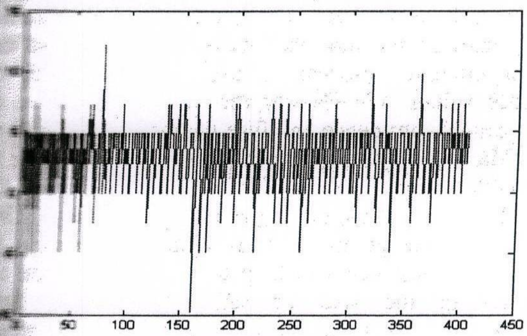

Fig.5. Light Sensor data

TABLE III. INDIVIDUAL ACCURACY RATE OF NAÏVE BAYES CLASSIFIER

Activity Type	1	2	3	avg.
Relaxing	0.8991	0.8273	0.5346	0.7537
Dish Washing	0.5494	0.5652	0.5124	0.5442
Meal preparation	0.5975	0.5975	0.6731	0.6227
Laundry	0.6440	0.6483	0.6935	0.6619
Cleaning	0.4798	0.4922	0.5519	0.5080

TABLE IV. INDIVIDUAL EFFICIENCYRATE OF NAÏVE BAYES CLASSIFIER

Activity Type	1	2	3	avg.
Relaxing	654.41	655.32	657.16	655.63
Dish Washing	650.68	657.68	654.50	654.29
Meal preparation	656.47	653.07	657.92	655.82
Laundry	654.83	658.76	657.31	656.97
Cleaning	659.75	658.83	656.64	658.40

In Table III and IV show the accuracy and efficiency of each category, respectively. One third portion of the dataset is used for testing and validation while seventy five percent of data have been used for the training purpose.

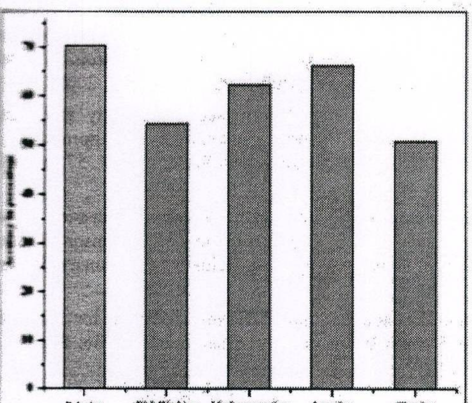

Fig.6. Accuracy of Naïve Bayes classifier

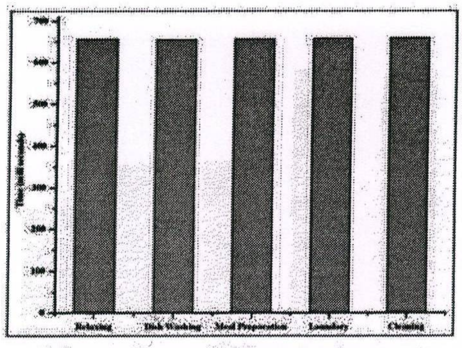

Fig.7. Efficiency of Naïve Bayes classifier

In the Fig. 6 and 7 accuracy and efficiency of each activity have been determined by applying the Naïve Bayes classifier for one hour data. In Multi-strategy approach, first discovers the relevant features on the basis of relevancy. Then inference process has been applied to build the confidence about the certain beliefs for improving the accuracy. Finally, max rule has been applied to recognize the activities. Fig. 8 and 9 show the accuracy and efficiency of each activity.

TABLE V. INDIVIDUAL ACCURACY RATE OF MULTI-STRATEGY APPRAOCH

Activity Type	1	2	3	avg.
Relaxing	0.7500	0.8000	0.7833	0.7782
Dish Washing	0.6833	0.6385	0.7258	0.6825
Meal preparation	0.7666	0.7166	0.7285	0.7372
Laundry	0.7400	0.7600	0.7200	0.7400
Cleaning	0.6666	0.6667	0.6500	0.6611

TABLE VI. INDIVIDUAL EFFICIENCYRATE OF MULTI-STRATEGY APPROACH

Activity Type	1	2	3	avg.
Relaxing	576.97	577.92	558.50	571.13
Dish Washing	580.65	576.76	577.65	578.31
Meal preparation	426.01	250.95	604.65	427.27
Laundry	423.33	423.31	423.39	423.49
Cleaning	350.63	347.51	349.56	349.23

Table V and VI show the accuracy and efficiency of each activity determined by applying the proposed Multi-Strategy Bayesian Model. One third portion of the dataset is used for testing and validation while seventy five percent data have been used for the training purpose.

Fig.8. Accuracy of Multi-Strategy Bayesian Model

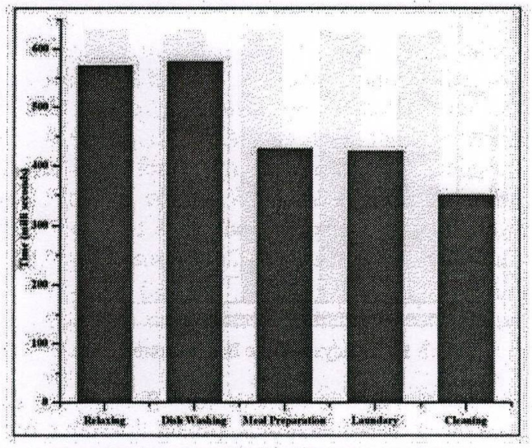

Fig.9. Efficiency of Multi-Strategy Bayesian Model

In the Fig. 8 and 9 accuracy and efficiency of each activity are determined by applying the proposed model for one hour data. We compare our Multi-strategy approach with the simple Naïve Bayes Classifier and get 9% higher accuracy.

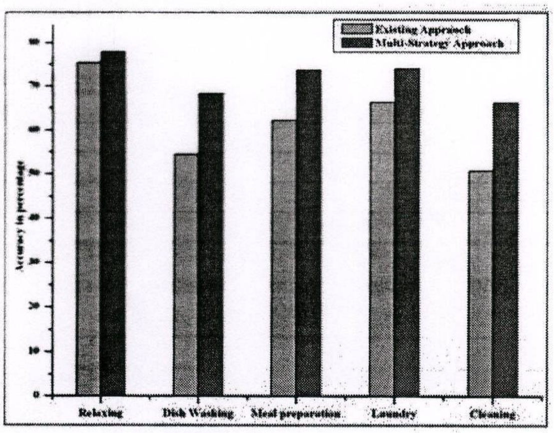

Fig.10. Comparison of Accuracy Rate

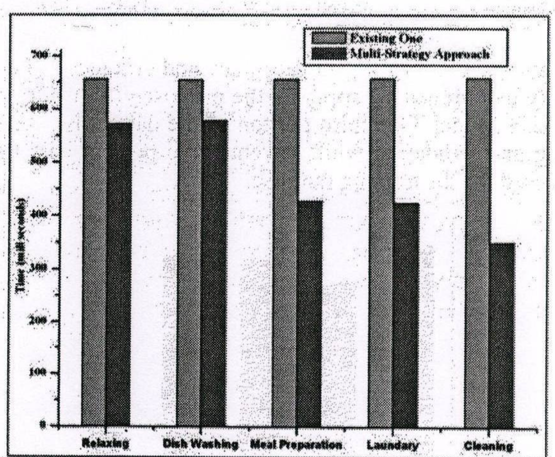

Fig.11. Comparison of Efficiency Rate

From Fig. 10 and 11, it is obvious that our approach is more accurate. Simple Naïve Bayes classifier is too much complex and time consuming. Our proposed method processes only selected features and fuses the individuals for identification of daily life activities in Smart Spaces.

V. Conclusion and Future Work

Our instance based learning technique K-nearest Neighbor reduced the complex search space by finding the relevancy in the feature vectors. Uniqueness of our idea is to get only relevant features with the help of kNN and then Bayesian inference process is used for calculating the confidence values. We classify the feature vectors on the basis of gained confidence and fuse the information with the help of Max rule. The experimental results demonstrate the considerable improvement in the execution time. Our proposed strategy is able to get the average accuracy 70.98% and 469 ms average time for recognizing the daily life activities. This test bed will help to cross validate the new techniques in the area of sensor fusion for Smart Environments. Our future work will focus on proposing stable frameworks and models for sensor fusion. We plan to explore and compare other fusion techniques that can be applied to sensor fusion in Smart environments.

Acknowledgment

The author would like to thank Dr. Donghai Guan and Mr. La The Vinh for their contributions in the improvement of algorithm and manuscript and providing materials related to this research.

This research was supported by the MKE (The Ministry of Knowledge Economy), Korea under the ITRC (Information Technology Research Center) support program supervised by the NIPA (National Industry Promotion Agency) (NIPA-2010-(C1090-1021-0003)).

References

[1] D. Lupiana, C. O'Driscoll and F. Mtenzi, "Taxonomy for Ubiquitous Computing Environments" IEEE Conference on Networked Digital Technologies, July 2009.

[2] G. D. Abowd and E. D. Mynatt, "Charting Past, Present, and Future Research in Ubiquitous Computing" ACM Transactions on Computer Human Interaction, Vol. 7,pp. 29–58, March 2000.

[3] F. Kobayashi, D. Masumoto and F. Kojima, "Sensor Selection based on Fuzzy Inference for Sensor Fusion", Proceedings of the IEEE International Conference on Fuzzy Systems, 2004

[4] H. Alex, M. Kumar and B. Shirazi, "Collaborating Agent Communities for Information Fusion and Decision Making" Knowledge Integration and Multi Agent, MA, USA, 2005.

[5] J. Sander and J. Beyerer, "Decreased Complexity and Increased Problem Specificity of Bayesian Fusion by Local Approaches" ACM journal of Robotics and Autonomous Systems, Vol. 57,pp. 259-267, March 2009.

[6] J. Sander, J. Beyerer,"Fusion agents- Realizing Bayesian fusion via Local approach" IEEE Conference on Multi-Sensor Fusion and Integration for Intelligent Systems,Heidelberg, Germany, September, 2006.

[7] W. Yuan-hong and L. Jun, "A New Method for Fish disease Diagnosis System Based on Rough Set and Classifier Fusion" IEEE

Conference on Artificial Intelligence and Computational Intelligence, 2009.

[8] S. S. Intille, K. Larson, E. M. Tapia,J. S. Beaudin, P. Kaushik, J. Nawyn and R. Rockinson"Using a Live-In Laboratory for Ubiquitous Computing Research" Proceedings of PERVASIVE, pp. 349-365, Berlin Heidelberg: Springer-Verlag, 2006.

[9] J. Sander and J. Beyerer, "A local approach for Bayesian fusion: Mathematical analysis and agent based conception",Journal of Robotics and Autonomous Systems,Vol. 57,pp. 259-267, 2009.

[10] O.R.Terrades, E.Valveny and S.Tabbone, "Optimal Classifier Fusion in a Non-Bayesian Probabilistic Framework", IEEE Transaction on Pattern Analysis and Machine Intelliegence,Vol. 31, pp. 1630-1644, September, 2009

[11] T. Michel, "Machine Learning" ISBN:0070428077, McGraw-Hill SciencePublisher, March 1997

[12] P. Cunningham and S. J. Delany,"K-Nearest Neighbour Classifiers" Technical Report University College Dublin,Mar 2007.

[13] R. P. W. Duin, "The Combining Classifier: to Train or Not to Train?" 16th IEEE International Conference on In Pattern Recognition, Vol.2, pp. 765-770, 2002.