Human Perception based Criminal Identification through Human Robot Interaction

Avinash Kumar Singh, Neha Baranwal and G C Nandi
Robotics and Artificial Intelligence Laboratory
Indian Institute of Information Technology, Allahabad, India
avinashkumarsingh1986@gmail.com, baranwal.neha24@gmail.com, gcnandi@gmail.com

***Abstract*—In India like many other countries, the crimes especially against women are rising. One of the many reasons perhaps is the low conviction rate. The initial stage of criminal investigation starts with the exploration of evidences and eyewitnesses. An eyewitness can act as a guide to trace out the suspect. Her/his perception about the suspect can be useful to identify the criminal. Based on the descriptions of the eyewitnesses normally a manual sketch is prepared and released in the Newspapers which most of the times is vague and ambiguous and hence ineffective. Therefore, a robotic framework has been proposed in this paper to help Police to identify criminals using the imprecise knowledge about the subject. The designed robotic system interacts with the eyewitness by asking several questions about the suspect such as age, height, her/his facial shape and size etc., and then making a guess about her/his face. A human face can be considered as the combination of various macro facial features such as eye. Eyebrows, face shape, lip shape, nose type etc. Although these features vary from person to person but their combination makes a human face unique. An experimental study on 113 Indian Bollywood celebrities, 37 Indian cricketers and 40 persons from Robotics and Artificial Intelligence Laboratory of Indian Institute of Technology Allahabad, India has been performed. The system is able to identify the criminal (here pseudo criminals) if it exists in the database. This system can be useful at the initial stage of investigation where we don't have any knowledge about the criminal.**

Keywords—Human Robot Intercation, Speech Recognition, Rough Set Theory, Artificial Intelligence

I. Introduction

In India like many other countries, the crimes specially against women are rising. Out of many reasons one might be low conviction rate. Sometimes the local police stations are not equipped with the modern criminal detection techniques. In this paper we propose to help law enforcing authorities by providing them a framework to identify the criminals through Human – Robot Interactions based on dialogue. Humanoid robots are apparently similar like human beings. Their human like design enable them to share the same workspace with humans. They can be used in various household applications (cleaning; cooking; washing etc.), sports (soccer, golf, etc.) and in assisting physically disabled persons. Apart from these their presence can also be seen in war field as mines lander in medical surgery using telerobotics. Here we propose a framework consisting of a humanoid robot capable of establishing the identity of a criminal through dialogue with the victims. It will be useful in those scenarios where an eyewitness has seen the suspect. The physical appearance of suspect is now in the mind of eyewitness. The proposed model described in next section will capture this knowledge about suspect and process it to find the match in the criminal database. Although there are

Various market available software [1][2] used by our police to generate artificial faces based on partial information of the suspect face, in all these cases we should have image of the suspect. But when we don't have the picture of the suspect then we usually use sketch drawer to draw the sketch of the Person. The sketch drawer embroid the imagination of eyewitness on the paper which is later being used to trace out the suspect. Our approach can be used in the initial phase before putting these things into action. The system works in the same philosophy of sketch drawer. The humanoid robot interacts with the eyewitness by asking questions about the suspect's age, height, gender and facial appearance etc. To establish this communication we have designed a speaker invariant speech recognition system which helps robot to understand the communication. The speaker invariant module makes this framework generalized so that it can be used for any victim. For demonstration a pseudo criminal database has also been developed for our proposed framework. It is a collection of photographs of Robotics Lab Staff and students together with some well recognized Bollywood celebrities and Indian cricketers. This database is further labelled in the presence of an expert who guided us to extract and label the database with different facial features [3] of a person. A rough set theory (RST) [4] based modeling has been added to make robot capable to predict the criminal. This prediction is purely based on the perception of the eyewitness and the availability of that criminal inside the database. The idea of using this system as the early stage of investigation is to reduce the work load of both the eyewitness and police. Imagine a scenario when we have 10 thousand faces in the criminal database. In this scenario eyewitness has to see each and every face to recognize the actual criminal. But by using this system automation, the search space is reduced to a very narrow domain which is very easier to work on.

This paper is organized as follows: In section II, we present a proposed framework, which includes Knowledge Representation; RST based modelling for criminal prediction. These modules are described in details in next sections III, IV respectively. Experimental validations of these modules are summarized in section V. This is followed by Section VI, where some conclusions and future work have been drawn.

II. PROPOSED FRAMEWORK

The proposed framework consists of four modules as shown in figure 1. Module 1 acts as an interface between robot and human. Robot interacts with the person by asking several questions about the appearance and looks of the suspect. Examples of knowledge capturing part (HRI) looks like this.

Robot: Can you tell me the gender (Male or Female)?
Human: Male.
Robot: Can you tell me his approximate age?
Human: 25 to 30
Robot: Can you tell me his approximate height?
Person: 5 foot 2 inch to 5 foot 8 inch.

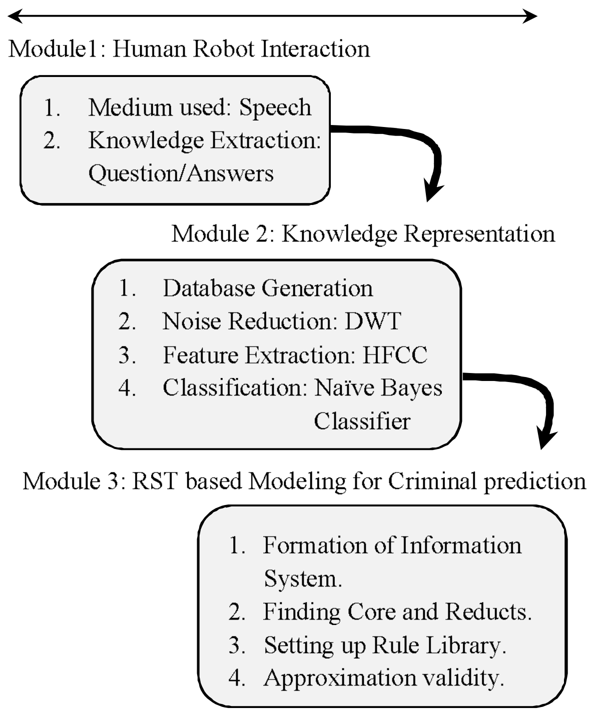

Figure 1: Proposed Framework

TABLE I. RAW DESCRIPTION OF THE ATTRIBUTE SET

Attribute / Feature	***Possible Values***	***Transformed Values***
Age (a_2)	BETWEEN var1 and var2	*
Gender (a_1)	Male/Female	M/F
Height (a_3)	BETWEEN var1 and var2	*
Face Shape (a_4)	Oval/Square/Round/Heart/Oblong	0/1/2/3/4/
Face Tone (a_5)	Very Light/Light/Light Brown/Brown/Black	0/1/2/3/4
Eye Shape (a_6)	Big/Normal/Small	0/1/2
Eyebrows Type (a_7)	Round/Flat/Hard Angle/Soft Angle/S Shape	0/1/2/3/4
Nose Type (a_8)	Long Thin/Long Wide/Small thin/Small Wide/Normal/Flat/Normal Wide/ Normal Flat	0/1/2/3/4/5/6/ 7
Lip Shape (a_9)	Big/Normal/Small	0/1/2

We have imposed some constraint over age and height to represent this information in symbolic form. Age is represented in years while height is represented in foot and inches. The Transformation is shown in table II.

Table II: Age and Height Transformation

	Symbolic Code								
Symbols	1	2	3	4	5	6	7	8	9
Age Constraint	15 20	21 25	26 30	31 35	36 40	41 45	46 50	51 55	56 60
Height Constraint	4.5 4.8	4.8 5.1	5.1 5.4	5.4 5.7	5.7 6.0	6.0 6.3	6.3 6.6		

There are total 9 questions about the face macro features are asked. Captured answers are processed though the speech processing module. These answers are further extended for feature representation as shown in table 1. The next 3rd and 4th module will process these features for predicting the accurate match inside the database. The prediction part is handled by Rough Set. The rough set model predicts the group/groups in which the suspect can belong.

III. KNOWLEDGE REPRESENTATION

This module takes the speech signal of human and recognizes the meaning of it [5]. This is basically a speech to text conversion unit which involves a proper training. The description of the training module and database used in this project is discussed in detail below.

Database Generation and Feature Extraction: We have created the database of isolated words listed in the second column of table I. Each word is recorded 10 times by 20 different peoples, 13 are males and 7 are females. The recording of each word is done using audacity software at 16000 Hz frequency of mono signals. Then DWT-4 is used for removing the noise present in the speech signal. In DWT [6] the audio signal is decomposed up to "L" level of approximate and detail coefficients where "L" is any integer number. The approximate and detail coefficients are obtained by using equation:

$$CD(j) = \sum_i s(i)g(2j - i)$$

$$CA(j) = \sum_i s(i)h(2j\text{-}i)$$

Where s(i) is the original speech signal, $1 \leq i \leq N$, where N is the number of sample values, CD(j) is the detail coefficients and CA(j) is the approximate coefficients calculated using high pass(g) and low pass(h) filter. The approximate coefficients are the low frequency components, which are highly sensitive in nature therefore it can be easily detected by human auditory system. The detail coefficients are the representation of the high frequency components, which are noisy in nature therefore it will be eliminate using low pass filter. In this paper we used DWT-4 wavelet transform where the length of filter bank is 4.

Here we have considered two types of features (MFCC, HFCC) are calculated from approximate coefficients of original audio signals. An HFCC [7][8] feature is a modification of MFCC features. The only difference is, in HFCC the filter bandwidth is independent of filter spacing due to central frequency fc property which is generated using Glasberg and Moore equation but in MFCC, vice versa. This property of HFCC provides independency towards speakers i.e. speaker invariant. The step by step explanation of extracting the HFCC coefficients from speech signals have been provided in [9].

Classification: Principal Component Analysis [10][11] has been used for further reducing dimensionality of extracted HFCC features. Classification is done using Bayesian classifier [12]. Here test speech signal is classified by calculating maximum probability density function as discussed below.

Bayesian Algorithm:

- Let there be C classes w1, w2 ….wC with prior probabilities P_1, P_2… P_C = 1/C and class conditional density functions: p_1, p_2,……p_C. Where 0<=P_i<=1 (for every i) and $\sum_{i=1}^{C} P_i = 1$.
- Reduced features obtained after PCA are considered as a training data set Y_i=(tr_1, tr_2……tr_m).
- Let X=(x_1, x_2, x_i) is a test vector put in a ith class. Where X ϵ π_i, π is a work space. $\Pi \subseteq |R^m$, x ϵ $|R^m$ set of all possible observations and m is the number of components.
- Here probability density function (PDF) is calculated using Gaussian multivariate distribution function is expressed as:

$$P_i(x) = argmax(\frac{1}{(\sqrt{2\pi})^m |\Sigma_i|^{\frac{1}{2}}} exp\left\{-\frac{1}{2}(X - \mu_i) \Sigma_i^{-1}(X - \mu_i)\right\}$$

Where Pi(X) is the PDF, μ=Mean of training data set, Σ=Co-variance of training data set, $|\Sigma|$=determinant of Co-variance, m is the dimension of given data. Test vector belongs to the class having highest PDF value. Here we consider C=12.

The information obtained from speech processing is further modelled using rough set theory (RST) for generating the actual face of the criminal. Speech processing module only gives the information about the parts of the criminal face not full face information. Therefore RST will be used for modelling of the criminal face.

IV. Rough Set Theory Based Criminal Prediction

The knowledge extracted from the eyewitness is imprecise. Therefore we have used several attributes to define proper feature type. The knowledge we captured is vague which could be processed by rough set theory [4][13]. RST is sufficient enough to handle the uncertainty and imprecision [14]. There are four basic steps involved in rough set modeling (1) Construction of a Decision Information System (2) calculating reducts and cores for the given information system (3) Generation of rules (4) Testing the validity of prediction, error analysis. All of these modules and key terms are discussed below.

A. Construction of Information System:

The information system (IS) is composed with the universe of discourse U and the attribute set A. It is represented by:

$$IS = (U, A)$$

The universe of discourse "U" consists of all the objects (U={o_1,o_2,o_3,o_4, … o_n }, here n is the number of objects ϵ U) and the attribute set is the set of all attributes required to represent these objects (A={a_1, a_2, a_3, a_4 …, a_m}, here m is the number of attributes). Each attribute could have several possible values which can be mapped by an attribute function($f_a: U \rightarrow V_a$), here $V_a$represents the value of particular attribute [15]. In our model the Information system consists of 190 classes and for each class 5 training views and 2 testing views exist. Therefore we have 950 objects and 9 attributes to describe them. These 9 attributes description was presented in Table–I. A transformation is required in order to encode these values so that it can be processed using the rough set theory. The transformation we used is shown in 3rd column of Table I.

For demonstrating our hypothesis we have randomly selected only 10 objects to represent the IS as shown in Table–III (please refer to Table-I for transformed values).

TABLE III : TRANSFOREMD INFORMATION SYSTEM

Objects	Conditional Attribute Set ($a_1 - a_9$), Decision Attribute (d)								
U	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	d
o_1	M	2	4	0	1	1	3	5	1	1
o_2	M	2	4	2	1	2	3	5	1	1
o_3	F	3	3	1	1	1	3	6	3	3
o_4	M	2	4	2	3	1	2	4	2	2
o_5	F	3	3	4	1	1	4	7	3	3
o_6	M	2	4	2	1	2	3	5	1	1
o_7	M	2	4	2	3	1	2	4	2	2
o_8	F	3	3	1	1	1	3	6	3	3
o_9	M	2	4	4	3	1	2	5	2	2
o_{10}	F	3	3	1	1	1	3	6	3	3
(950 Samples from the Criminal Dataset, 190 classes)										

The kernel of RST is to discover the indiscernible relation out of the given information system. The indiscernible relation is defined on two objects o_i and o_j where $i \neq j$ as: $a(o_i) = a(o_j)$ for every $a \in A$. Object that belongs to same indiscernible relation is represented as $IND(A)$ known as elementary sets and it is denoted by$[o_i]_{IND(A)}$[15]. The example of elementary sets created on attribute set A is shown in table –IV.

TABLE IV: ELEMENTARY SET DESCRIPTION OF "IS"

U/A	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	d
$\{o_3, o_8, o_{10}\}$	F	3	3	1	1	1	3	6	3	3
$\{o_2, o_6\}$	M	2	4	2	3	1	2	4	2	1
$\{o_4, o_7\}$	M	2	4	2	1	2	3	5	1	2
$\{o_1\}$	M	2	4	0	1	1	3	5	1	1
$\{o_5\}$	F	3	3	4	1	1	4	7	3	3
$\{o_9\}$	M	2	4	4	3	1	2	5	2	2

B. *Finding Core and Reducts:*

The attribute used to represent the system is further processed and analyzed in this section. The motive of finding core and reducts of the "IS" is to find the independent set of attributes. Reducts of the attribute set are the minimal subset which leads to the same partitions (elementary sets) as we achieved on the full set. These attributes will later help in setup the rules library. Core shows the presence of single attribute which discerns from other attribute of the other object [15]. But before proceeding for core and reducts some key ingredients which will be useful in this section should be discussed.

Lower, Upper and Boundary Approximation: There could be three condition arises for predicting the data over one of the elementary set which will be expressed by these three categories.

i. Lower Approximation: Elements possibly belongs to the elementary set.
ii. Upper Approximation: Elements those certainly belongs to the elementary set.
iii. Boundary Approximation: Elements those belong to the boundary sets.

Let B is the subset of U ($B \subset U$) then the lower approximation of B is defined by the union of all the elementary sets which fall into the domain of B. The lower approximation is denoted by $\underline{BO}$.

$$\underline{BO} = \{ o_i \epsilon U \mid [o_i]_{IND(B)} \subset O\}$$

Similarly the upper approximation of B is defined as:

$$BO = \{ o_i \epsilon U \mid [o_i]_{IND(B)} \cap O \neq \emptyset\}$$

Any object which belongs to the lower set can also be belongs to the upper set. The boundary approximation is defined by the intersection of lower and upper approximation. It can be represented as: $BNO = BO - \underline{BO}$

Example: Let B is a group of some objects B= $\{o_2, o_3, o_4, o_6, o_7\}$. Then the lower approximation set will be defined as: Elementary set: $\{o_2, o_6\}$ and $\{o_4, o_7\}$

$$\underline{BO} = \{ o_2, o_4, o_6, o_7 \}$$

The upper approximation set will consider the common attribute of $\underline{BO}$ also.

Elementary set: $\{o_2, o_6\}$, $\{o_4, o_7\}$ and $\{o_3, o_8, o_{10}\}$

$$BO = \{ o_2, o_3, o_4, o_6, o_7, o_8, o_{10} \}$$

The boundary approximation will be calculated as:

$$BNO = \{ o_2, o_3, o_4, o_6, o_7, o_8, o_{10} \} - \{ o_2, o_4, o_6, o_7 \} = \{ o_3, o_8, o_{10}\}$$

Approximation Accuracy: Accuracy of these approximations is also represented by rough membership and it is calculated using:

$$\mu_B(O) = card(\underline{BO})/crad(BO)$$

Here $card(\underline{BO})$ represents the cardinality of lower approximation which is basically the number of elements present in the lower approximation elementary set. Similarly the $card(BO)$ represents the elements present in the upper approximation elementary set.

Here $card(\underline{BO}) = 4$ and $card(BO) = 7$

Then $Accuracy$: $\mu_B(O) = \frac{4}{7} = 0.57$

Computing Core and Reducts: Core and Reducts are computed with the help of descernibility matrix. The discernibility matrix is square matrix (nXn) where n is the number of elementary sets. The discernible relation of two objects $o_i$and $o_j$shows how they discern to each other based on their attribute set. An example is demonstrated in Table IV. The set represented in the table are:

TABLE V: DESCERNIBLITY MATRIX

	S_1	S_2	S_3	S_4	S_5	S_6
S_1	-	$a_1, a_2, a_3, a_4, a_5, a_7, a_8, a_9$	$a_1, a_2, a_3, a_4, a_6, a_8, a_9$	$a_1, a_2, a_3, a_4, a_8, a_9$	a_4, a_7, a_8	$a_1, a_2, a_3, a_4, a_5, a_7, a_8, a_9$
S_2		-	a_5, a_6, a_7, a_8, a_9	a_5, a_4, a_7, a_8, a_9	$a_1, a_2, a_3, a_4, a_5, a_7, a_8, a_9$	a_8
S_3			-	a_4	$a_1, a_2, a_3, a_4, a_6, a_7, a_8, a_9$	a_4, a_5, a_7, a_6, a_9
S_4				-	$a_1, a_2, a_3, a_4, a_7, a_8, a_9$	a_4, a_5, a_7, a_9
S_5					-	$a_1, a_2, a_3, a_5, a_7, a_8, a_9$
S_6						-

Set 1(S_1): $\{o_3, o_8, o_{10}\}$. Set 2 (S_2): $\{o_2, o_6\}$. Set 3 (S_3): $\{o_4, o_7\}$. Set 4 (S_4): $\{o_1\}$. Set 5 (S_5): $\{o_5\}$. Set 6 (S_6): $\{o_9\}$

The discerrniblity function $f(A)$ will be applied to constructs the reducts for the calculated discernible matrix. Here we have used 5 descriniblity functions for each elementary set. Like $f_1(A)$take care only about how the elementary set 1 differs from 2, 3, 4, 5 and 6.

$$f_1(A) = (a_1 + a_2 + a_4 + a_3 + a_5 + a_7 + a_8 + a_9) X (a_1 + a_2 + a_4 + a_3 + a_6 + a_7 + a_8 + a_9) X (a_1 + a_2 + a_3 + a_5 + a_8 + a_9) X (a_4 + a_7 + a_8) X (a_1 + a_2 + a_4 + a_3 + a_5 + a_7 + a_8 + a_9)$$

$$= (a_4 + a_7 + a_8)$$

$$f_2(A) = (a_5 + a_6 + a_7 + a_8 + a_9)\, X\, (a_5 + a_4 + a_7 + a_8 + a_9)\, X\, (a_1 + a_2 + a_4 + a_3 + a_5 + a_7 + a_8 + a_9)\, X\, (a_8)$$

$$= a_5a_8 + a_7a_8 + a_8a_9 + a_8$$

Similarly we can calculate the discernibility function $f_3(A)$, $f_4(A)$ and $f_5(A)$. There are 11 different reducts we achieved from each function. These are the set of optimal attribute which will result the same partition. We can design our rules on the basis of these optimal attribute sets.

C. Decision Rules:

Decision rules are constructed based on the decision table. A decision table is same as the Information system with the additional column decision (d). The decision table is shown in Table IV. The decision rules are created based on these elementary sets shown below.

$$a_{1_F} \wedge a_{2_3} \wedge a_{3_3} \wedge a_{4_1} \wedge a_{5_1} \wedge a_{6_1} \wedge a_{7_3} \wedge a_{8_6} \wedge a_{9_3} \Rightarrow C_3$$
$$a_{1_M} \wedge a_{2_2} \wedge a_{3_4} \wedge a_{4_2} \wedge a_{5_3} \wedge a_{6_1} \wedge a_{7_2} \wedge a_{8_4} \wedge a_{9_2} \Rightarrow C_1$$
$$a_{1_M} \wedge a_{2_2} \wedge a_{3_4} \wedge a_{4_2} \wedge a_{5_1} \wedge a_{6_1} \wedge a_{7_3} \wedge a_{8_5} \wedge a_{9_1} \Rightarrow C_2$$
$$a_{1_M} \wedge a_{2_2} \wedge a_{3_4} \wedge a_{4_0} \wedge a_{5_1} \wedge a_{6_1} \wedge a_{7_3} \wedge a_{8_5} \wedge a_{9_1} \Rightarrow C_1$$
$$a_{1_F} \wedge a_{2_3} \wedge a_{3_3} \wedge a_{4_4} \wedge a_{5_1} \wedge a_{6_1} \wedge a_{7_4} \wedge a_{8_7} \wedge a_{9_3} \Rightarrow C_3$$
$$a_{1_M} \wedge a_{2_2} \wedge a_{3_4} \wedge a_{4_4} \wedge a_{5_3} \wedge a_{6_1} \wedge a_{7_2} \wedge a_{8_5} \wedge a_{9_2} \Rightarrow C_2$$

Rules generated from above tables are connected with the conjunction or disjunction operators shown below. Here a_{k_i}represents that the attribute "k" has "i" value (k[1,11])
The elementary set described in table IV is the basic one. The rules In order to optimize the search we have discovered the indiscernible relation. Let see another scenario shown in table on the reduct set $\{a_8a_9\}$ generated in function f2 Table VI.

TABLE VI: DECISION TABLE FOR REDUCT SET $\{a_8a_9\}$

U	a_8	a_9	d
o_1	5	1	Criminal 1
o_2	5	1	Criminal 1
o_3	6	3	Criminal 3
o_4	4	2	Criminal 2
o_5	7	3	Criminal 3
o_6	5	1	Criminal 1
o_7	4	2	Criminal 2
o_8	6	3	Criminal 3
o_9	5	2	Criminal 2
o_{10}	6	3	Criminal 3

The following rules can be extracted from Table VI.

If a_8=5 and a_9=1 => d = Criminal 1
If a_8=4 and a_9=2 => d = Criminal 2
If a_8=6 and a_9=3 => d = Criminal 3

V. EXPERIMENTAL VALIDATION

We have categorized this section into two modules. These two modules are different in their principle and working category, so it is better to provide the proof of each technique separately.

COMMUNICATION RELIABILITY:
The reliability of communication between robot and human is purely based on the classification accuracy of speech signal. In order to test this communication we have created training and test set offline. The training set consists of all the words which is discussed in second column of the Table-I. The details of training and test set specification are provided below together with their accuracy in Table VII. In order to perform the entire experiment, we have used Logitech H150 microphone as an input device and captured the test sound signals at the distance of 5, 10 and 15 meters. Total 20 persons were used in the training phase. 55 unique isolated words used to capture the imprecise knowledge of eyewitness with 10 samples of each word to train the system. The system is tested in two environment condition (a) Silence (b) Lab Environment.

TABLE VII: CLASSIFICATION RESULTS UNDER FDIFFERENT CONDITIONS

Distance	5 meters	10 meters	15 meters
Environment	Classification Accuracy (%)		
Silence	98	97	97
Laboratory	95	89	85

From Table–VII we can conclude that as the distance between human and robot increases the classification accuracy decreases. The classification accuracy of HFCC based features confirms the applicability of the system in human robot interaction. The system is independent of the speaker; hence it will work for all kind of eyewitnesses. All the information obtained from speech recognition module is stored in text format so that it will be used in RST module.

EXPERIMENTAL VALIDATION OF RST BASED PREDICTION:
The speech recognition unit feeds all nine features to the RST module. RST utilizes these features to predict the actual class labels. We have used all 190 criminal faces (150 celebrities and 40 Robita face database) to model the system with 9 attributes to represent the IS. We have achieved 64 elementary sets after applying RST. These elementary sets are considered as class if any singleton elementary sets exits we group them to form a class. Almost all elementary sets possess equal number of objects in their lower and upper approximations. Therefore all are having equal membership value as 1. We have also achieved 64 reducts having different attributes in there attribute sets. A decision table is also being formed for each reducts which guided us to setup the rule library. The list of reducts with set of rules achieved through each function is summarized here in table VIII.

TABLE VIII: SET OF REDUCTS WITH RULES

Reducts	Rules	Reducts	Rules
(a_4, a_7, a_8)	6	(a_4, a_9)	5
(a_5, a_8)	5	(a_4, a_7, a_9)	6
(a_7, a_8)	5	$(a_1, a_2, a_3, a_5, a_7, a_8, a_9)$	5
(a_8, a_9)	5	(a_4, a_7)	5
(a_4, a_6)	5	(a_4, a_8, a_9)	5

Almost all reducts are giving a same level of accuracy of 92.5%. For testing the validity of our modelled system we have generated 380 random records those are similar like training having only little deviation and predicted the class level for each.

EXPERIMENTAL VALIDATION OF COMPLETE FRAMEWORK: The presented experiment is executed and tested on our Robotics and artificial intelligence lab (ROBITA LAB) made SMART robot shown in figure 2. SMART is designed to assist guests in ROBITA LAB demonstration. It has several functionality associated with it like it can mimic some gestures e.g shake hand, Namaste (greeting), hi etc. Smart is equipped with wheels which enables it to move throughout the defined workspace. A camera is mounted at the top used to guide it while the microphone helps it to pursue natural languages (human speech). All these peripheral devices are attached with the laptop having facility to perform several modules like speech recognition, computer vision, microcontroller logics etc. This setup is used for establishing the proof of proposed framework. We have used 20 criminal faces and 20 eyewitnesses associated with each criminal. Each eyewitness has seen the photograph of the criminal in detail. Later, questions related to physical and facial appearance of the criminals were asked by SMART to each eyewitness. All the responses are processed through SMART's speech recognition unit and converted to its corresponding classes. The collections of these responses are named as attributes which is input to the RST based prediction. We have already defined some rule library in the training process of RST. Based on these attributes some rules are fired from rule library and mapped these attributes to any of the existing classes (Please refer to section 3). It is possible that these rules can predict different classes for that particular test attribute set. The final decision is summarized in terms of 10 most matching criminal faces. We have used the span of 10 because most of the time the criminal face fall in top 10 matched records. In conclusion we saw that if speech recognition module of smart is misclassifying at some attributes than this imprecision and uncertainties are handled by the RST.

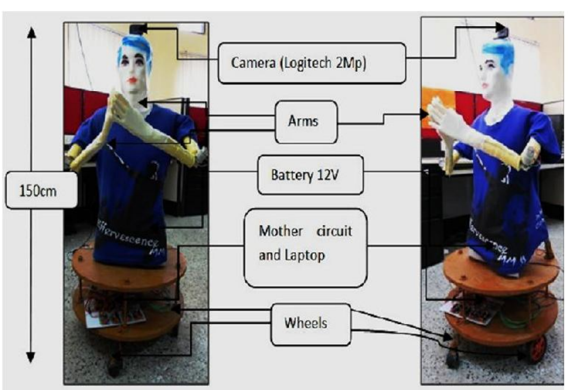

Figure 2: Social Mobile Advanced Robot Test-Bed (SMART)

VI. CONCLUSION AND FUTURE WORK

An effort has been put in this paper to make humanoid robot framework capable to identify criminals through dialog with the victims to help the law enforcement authority. This designed framework has been tested on 113 Indian Bollywood celebrities, 37 Indian cricketers and on 40 persons of our IIIT-A, Robotics & Artificial Intelligence Lab. Modeling of the problems has been done using Rough Set Theory (RST). The RST helps us to find the minimal subset of the attribute set while keeping the maximum prediction rate. We have achieved prediction accuracy of 92.5%. Here RST based system is helping to reduce the search space. The search span has been minimized up to 10 which mean out 190 records we will see only 10 photos of the criminal. This reduces human effort and speedup the process of recognition.

Accuracy of approach is highly dependent on the perception of eyewitness. Therefore, we have modelled both the systems in such a way that they can handle the vague perception of eyewitness. But it is desirable that the perception should be nearer to the reality.

REFERENCES

[1] http://www.facesid.com/products_faces_le.html, visited on 5 July 2013.

[2] http://animetrics.com/forensicagps, visited on 5 July 2013.

[3] http://design.tutsplus.com/tutorials/human-anatomy-fundamentals-advanced-facial-features--cms-20683, visited on 5 July 2014.

[4] Z. Pawlak, Rough sets, Int. J. Inf. Comput. Sci. 11 (1982), pp. 341–356.

[5] Xuedong Huang and Kai -Fu Lee. On Speaker-Independent, Speaker-Dependent, and Speaker- Adaptive Speech Recognition. In Proceedings of IEEE Trans. on speech and audio processing. vol.1, no.2, April, 1993.

[6] C. Lin and Liu. A Tutorial of Wavelet Transform. February 23, 2010.

[7] Robert Wielgat, Tomasz P. Zielinski, Tomasz Potempa, Agnieszka Lisowska-Lis and Daniel Korl. HFCC based recognition of bird species. In Proceedings of IEEE Signal Processing Algorithms, Architectures, Arrangements and Applications, sept, 7, 2007.

[8] Mark D. Skowronski and John G. Harris. Human factor cepestral coefficients. First Pan-American/Liberian meeting on acoustics primera reunion pan-Americana/iberica de acoustical, dec 2-6. 2002.

[9] Baranwal, Neha, Shweta Tripathi, and G. C. Nandi. "A speaker invariant speech recognition technique using HFCC features in isolated Hindi words."International Journal of Computational Intelligence Studies 3.4 (2014): 277-291.

[10] Singh, Avinash Kumar, Pankaj Joshi, and Gora Chand Nandi. "Face recognition with liveness detection using eye and mouth movement." Signal Propagation and Computer Technology (ICSPCT), 2014 International Conference on. IEEE, 2014.

[11] Singh, Avinash Kumar, and G. C. Nandi. "Face recognition using facial symmetry." Proceedings of the Second International Conference on Computational Science, Engineering and Information Technology. ACM, 2012.

[12] Neri Merhav and Yariv Ephraim. A Bayesian classification approach with application to speech recognition. IEEE transactions on signal processing. Vol.39, No. 10, October 1999.

[13] Z. Pawlak, Rough Sets. Theoretical Aspects of Reasoning about Data, Kluwer Academic Publisher, Dordrecht, Netherlands,1991.

[14] Ivo Düntsch, Günther Gediga, Uncertainty measures of rough set prediction, Artificial Intelligence, Volume 106, Issue 1, November 1998, Pages 109-137

[15] Walczak, B., & Massart, D. L. (1999). Rough sets theory. Chemometrics and Intelligent Laboratory Systems, 47 (1), 1–16.