Int. J. Mach. Learn. & Cyber. (2019) 10:413–431
DOI 10.1007/s13042-017-0699-z

ORIGINAL ARTICLE

A rough set based reasoning approach for criminal identification

Avinash Kumar Singh[1] · Neha Baranwal · G. C. Nandi[1]

Received: 28 October 2015 / Accepted: 29 April 2017 / Published online: 30 June 2017

Abstract As a supplement to mugshot detection, a new approach is proposed to capture the eyewitness's visual perception in the form of symbolic representation. It reveals physiological and facial characteristics of criminal which help in their identification. A rough set theory based technique is introduced to model those symbolic representations. This approach provides an intuitive insight to process criminal's imprecise and imperfect knowledge. We used a benchmark mug-shot dataset consisting of 300 criminals faces from the Chinese University of Hong Kong (CUHK) to study the correctness of our proposed model. We took the help of 105 students of Indian Institute of Information Technology, Allahabad, who were treated as eyewitness to depict the visual perception about 300 criminal faces of CUHK. The experimental verification is composed of two modes which are analogous to viewed sketches and forensic sketches. Like viewed sketches we have generated test case-I, where perception is given while looking at the photo whereas test case-II is like the forensic sketches where the description is given by recalling the memory. We have achieved encouraging results on the viewed sketch database as well as forensic sketch database.

✉ Avinash Kumar Singh
avinashkumarsingh1986@gmail.com

Neha Baranwal
baranwal.neha24@gmail.com

G. C. Nandi
gcnandi@iiita.ac.in

[1] Robotics and Artificial Intelligence Laboratory, Indian Institute of Information Technology Allahabad, Allahabad, UP 211012, India

Keywords Criminal identification · Rough set theory · Dimensionality reduction · Classification · Information processing

1 Introduction

A person can be categorized from his/her behavioural, physiological and facial attributes or the combination of all three [1]. Person's behaviour can be captured in terms of his/her gait patterns and the way of speaking, his/her physic can be represented in the form of age, gender and height, while the face description can be summarized as face category, face tone, eyebrows type, eye shape, nose shape and lip size. This paper utilizes all these attributes and talks about their contribution to forensic domain more precisely on criminal identification. Evidence collection and find a reliable match for the collected evidences are the two major blocks of forensic science [2]. Aforesaid attributes are served as evidences to trace the criminal, specially facial information about the suspect plays a vital role in criminal identification. The previous summary of the suspect can be filtered out just by matching the photograph collected from the digital sources in the existing criminal or citizen face database. Sometimes due to the absence of digital media such as CCTV or camera, the face information of the suspect is not available. In all such cases, an eyewitness who has seen the crime and the suspect can provide some description. In all these cases, a police sketch artist communicates with the eyewitness and draws a portrait of the suspect based on the given description [3, 4]. The problem of manual sketch creation is trivial and requires the human expertise as well as large amount of time.

Although there are commercial softwares available in the market [5–7] which can reduce the effort and produce the criminal sketch. But, their accuracy and reliability depend on how good the proper features are selected and fine-tuned. On the other hand, once the sketch is ready it is released in the public media and sent to all nearby police stations. The process of finding the criminal in the existing criminal photo album is a tedious task. In the past research, researchers have designed an automatic way of finding the match. This procedure is known as mug-shot detection, which processes the sketch image to find the match from the criminal face database. The first step towards the sketch to photo matching was put by Roberg et al. [8]. The experiment was performed on small database of 7 sketches and 17 real photographs, however, they managed to prove that this process could be further enhanced. Later the significant contribution in this direction was made by [9–11]. Unfortunately, this problem has not received proper attention and requires further research in this direction.

The existing solutions work only in constraint environment where the gap between the sketch and photo is less. On the other hand they have utilized only facial features to trace out the suspect. They did not utilize the other two features (behavioural and physiological). The addition of these two additional attributes can improve the recognition accuracy [12]. These issues motivated us to look for other options which could be useful. If we closely analyse the process of sketch embroidery, we can find these sketches are merely representation of imprecise knowledge of the eyewitness. The same can also be recorded by verbal communication with eyewitness. This paper presents a prototype of each facial attribute which is helpful to capture the perception about a suspect. There are basically three different categories of these facial features discussed in the literature [13], viz. visual, anthropomorphic and cephalometric features. The prototype library is set up after analyzing these categories and features. A semi-automated (requiring expert guidance) system is proposed in this paper to process the imprecise knowledge of the eyewitness and to obtain desired matching results from the existing database. A framework is designed which consists of 300 criminal faces (100 from CUHK dataset + 100 from AR dataset + 100 from FERET dataset). We invited 105 volunteers to give the description about these faces. Each volunteer contributed his/her perception for 20 criminals using the interface depicted in Fig. 1. Each criminal is summarized in terms of seven views (five training views and two test views). We designed an expert system based on these training views using the rough set theory to predict the criminal. The prediction is summarized in terms of top ten most matching faces arranged in descending order. Seeing the capability of the system, it can be used before mugshot module and avoid the process of sketch creation. Therefore, the proposed system acts as a supplementary to the mugshot detection, not the supplant.

The rest of the paper is organized as follows: Sect. 2, presents the state of the art on sketch to photo matching, followed by Sect. 3 which describes the design issues of the system. Section 4 summarizes the background of rough set and case based reasoning while Sect. 5 defines their implementation on this issue. At the end Sect. 6 presents proof of our hypothesis in terms of results and validations. Finally Sect. 7 concludes the paper with advantages, limitation and future scope of this new way of criminal identification.

2 State of the art

The work in mugshot detection was started by Roberg et al. [8] who has created a system which gave the predicted possible matches based on the presented sketch images. The matching between the gallery (photo images) and the probe images (sketches) were based on the aggregated matching score of full face, cropped face and patches of the face. Two

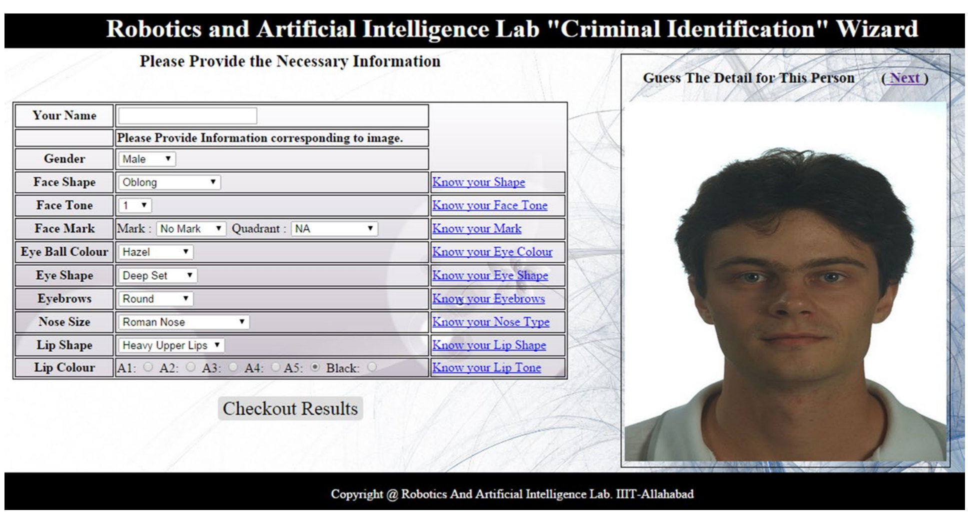


Fig. 1 Designed system (user interface)

normalization techniques, geometric standardization and photometric standardization were used for photo as well as sketch images. The geometric standardization used CANDID model to normalize photo images with respect to their respective sketch images, while the photometric standardization transformed input photograph to pseudo photograph. These pseudo photographs represented the texture information of the sketch images in the form of background and could remove high intensity regions which can cause blur effect on these images. The system was tested on the very low datasets 7 sketches and 16 real photographs.

Another work in this domain in 2003 was of Tang et al. [14], who introduced the concept of Eigensketches in order to represent the photo images. They provided the proof on 188 photo/sketch images (88 training and 100 testing) that these Eigensketches minimized the gap between photo and sketch images to render better classification. The Eigensketch concept is pretty much similar as Eigenfaces approach of Turk and Pentland [8], however they differ in the sense of their final projection of input face images. The photo images were transformed into the Eigensketches using the Karhunen Loeve Transform (KLT). This transformation allows them to represent the input photo image into the linear combination of other training photo images. This brings out structural information of the photo image which is very effective in finding the suitable match.

From 1996 to 2002, the isolated progress was seen in the field of artificial sketch designing [5–7]. Researchers were able to generate these sketch images directly from the input photo images. In these cases, there was no need of a sketch drawer to prepare a sketch manually. Later in 2004, Tang and Wang [15] came with their modified version of Eigensketches [14]. This time they introduced the notion of shape and texture based Eigensketches. According to Tang et al. shape and texture are two distinguished elements with respect to matching between sketches and real photographs. Shape is used by the sketch drawer to give importance to the particular facial component while texture is used to portrait 3D structure (ups and downs) of the face. These two main features are extracted using facial fiducial points and mean face shape of the training images. Eigen transformation is applied on these extracted shape and textures information of photo images. These transformed images are used as input for Principal Component Analysis (PCA) and Bayes classifier to classify to one of the classes. In continuation of their previous approaches [14, 15], Tang and Wang [9] used Eigenspace transformation to minimize the gap between the sketch and photo images. Here they tested their hypothesis on 188 population of photo and sketch images. As in the previous approach and henceforth they transformed the input face images to pseudo sketch images and further transformed raw sketch images to transformed Eigensketches. The coefficient of both the transformed spaces are used to find out the distance between gallery and probe. Testing on three different distance parameters shows the efficiency of their approach. He has achieved around 96% accuracy and proved that it is better in some scenarios than the human.

It is well accepted truth that whenever an artist designs the sketches, it is possible that there would be distortions. These distortions disturb the solution and make the problem nonlinear in nature. Initially this issue has not been addressed by [8, 11, 14]. It prompted Liu et al. [16] to work in this direction. They introduced the concept of local geometry preserving nonlinear transformation function to handle this non linearity. This transformation is motivated by Local Linear embedding and helps in generating pseudo sketch images. The non-linearity is also handled by using the kernel based nonlinear discriminant analysis. The same problem of non-linearity is also solved by Embedded Hidden Markov Model (E-HMM) based facial sketch synthesis (FSS) by Gao et al. [13]. They used the same theory of wang to generate the pseudo Eigensketehes but using different HMMs. They assumed that the transformation from photo images to pseudo is nonlinear hence the nonlinear transformation is required. This nonlinear transformation is modelled with E-HMM. Later the modified version of this approach has been seen in [11, 17]. Although these approaches addressed the problem of non-linearity effectively but they produce blurring effect in pseudo sketches and photos. It has also been noticed that these techniques are producing block edges as effect. This problem is further investigated by Xiao et al. [11] by introducing the notion of subspace learning to generate synthesized photo sketches.

The previous literature had been studying the problem of sketch to actual photo prediction based on the Eigenspace transformation but recently in Klare et al. [10] have enlightened the subject by new ways of recognition. They have used the fundamental technique feature extraction in this field in the form of local feature based discriminant analysis (LFDA), multi scale local binary patterns (MLBP) and scale invariant feature transform (SIFT). Here the SIFT and MLBP are used to extract the local features from both the sketch images and the real photographs, while LFDA is applied on these local descriptors to refine and reduce the dimensionality. Further, the matching is performed in the transformed feature space. There are two kinds of scenarios addressed in the previous literature. In the first case the sketch is prepared by the photo given to the sketcher, while in second case only description about the face is available. Most of the literature address the first type of issue. Both of these cases are addressed by [10]. They tested the validity of their proposed approach and figured out that forensic sketches (second case) are affected by the biased perception of eyewitness. Therefore good results have not been obtained so far over viewed sketches (first case). This problem has not received proper attention and requires further research in this direction. The existing solutions help only for constraint environment where the gap between the sketch and photo is less. These issues motivated us to look for other options which could be useful. If we closely analyse the process of sketch embroidery, we can find that these sketches are the representations of imprecise knowledge of eyewitness. The same can also be recorded by verbal communication with eyewitness.

The other field of literature describes facial attributes for criminal identification [18–20]. Recently Klare et al. [19] presented an attribute based face recognition system which is helpful in criminal identification. They have considered 46 major attributes for defining the face and facial features. The attribute set consists of features such as lip thickness, nose size, face shape, eye colour, eyebrow position, hair colour each. These attributes are further e matching and retrievaxpressed in terms of holistic feature (such as gender or wrinkles), component level information (eyes, nose, mouth etc) and the relationship between components (distance between the nose and mouth). There are 19 features which are composed by binary values (unibrows (when two eyebrows are connected)) and no unibrow), 19 features have three category values (nose size small, normal and large), 6 attributes have four category values and other two attributes have 6 and 7 category values. The category value for each attribute is used to quantify that attribute. The other approach analyses semantic knowledge of face images to retrieve culprit photograph from the face database Gudivada et al. [18]. The semantic knowledge from the face images are extracted based on the personal construction theory. The semantic of the attributes are defined based on the attributes subjectivity, imprecision and uncertainty. Subjectivity is due to the different view point. The difficulty in measurement and specification leads to imprecision and uncertainty. The decision making of any person could be influenced by the environment/objects. Other researchers have used multi attribute for image ranking and image retrieval Siddiquie et al. [20]. They demonstrated the automatic extraction and labelling of different facial and anatomical features of a person. There are 27 features used to represent the person information. These features include the origin (nationality), face, hair and eyeball colour, gender etc. They have also considered the interdependency between attributes in retrieving information from the database.

3 Knowledge acquisition and representation

3.1 Knowledge acquisition

Knowledge acquisition is the first step towards designing the expert system. In order to extract valuable knowledge from the user we created a web interface. The web interface asks various queries about the criminal's physical and facial attributes. A snap shot of the designed system is shown here in Fig. 1.

Although every human face has facial attributes like eyes, nose, mouth etc. They all are placed at the same place and position with respect to every face. It is also possible that two persons can have same kind of nose or eyes (genetically this is gifted by father or mother to their children). But their varying placements make the human face unique thereby we can differentiate one from other. We tried to extract the possible feature description through our web interface. The feature description is presented one by one here.

3.1.1 Face type

Based on face anthropometric structure [21] a face can be categorized over one of the classes shown in Fig. 2, based on its outer boundary.

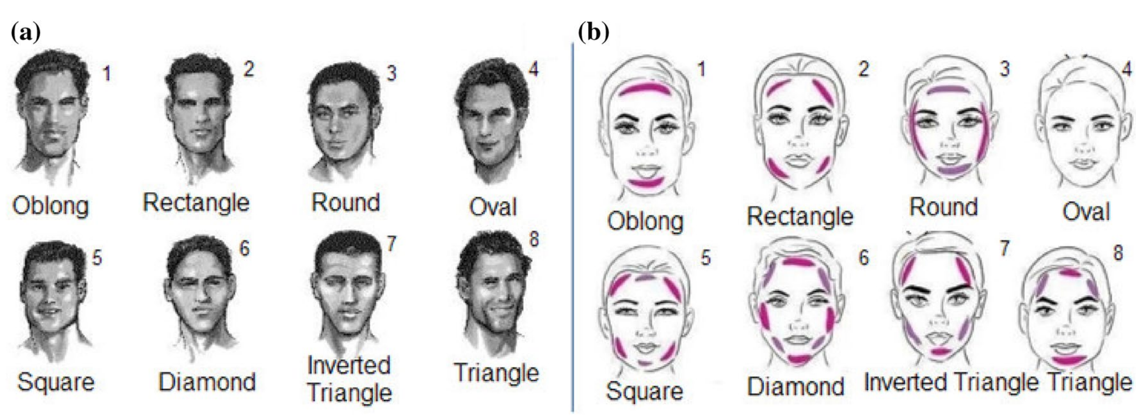


Fig. 2 Different shapes of face [10]

3.1.2 Face tone

Face tone speaks several things about your origin, your geographical origin etc. [22]. It can be described by a face tone parameter presented below in Fig. 3.

3.1.3 Eye shape

As like the face tone and shape type Eye has also different kinds of shape, described below in Fig. 4 [23].

1. **Deep set eyes** Deep set eyes are usually bigger in size and set as deeper into skull causes more salient brow bone.
2. **Monolid** Monolid kind of eyes represents flat with respect to facial anatomy. No stronger crease and not very much prominent brow bones visibility.
3. **Hooded eyes** Brow is covered with extra layer of skin resultant upper eye lids looks smaller.
4. **Protruding eyes** Eye lids seem to be bulging from the eye socket. Such eye lids make the eyes attractive.
5. **Upturned eyes** They look like an almond shape eye with the kick in outer corner.
6. **Downturned eyes** They are same in shape as of upturned eyes only a drop in outer corner.
7. **Close set eyes** When the eye ball distance between two eyes is minimum.
8. **Wide set eyes** Just opposite the close set eyes, here the distance is more between eye balls.

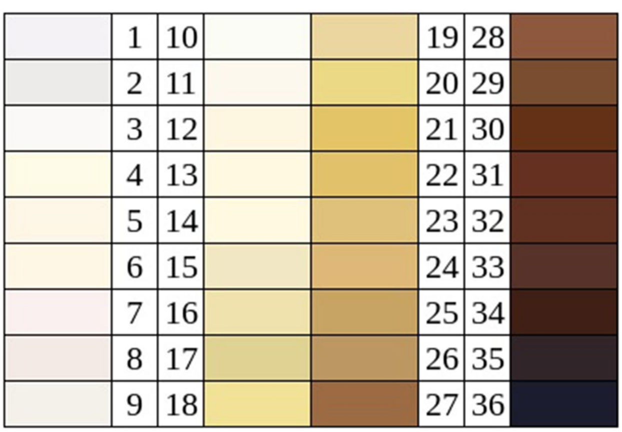


Fig. 3 Face skin tone [22]

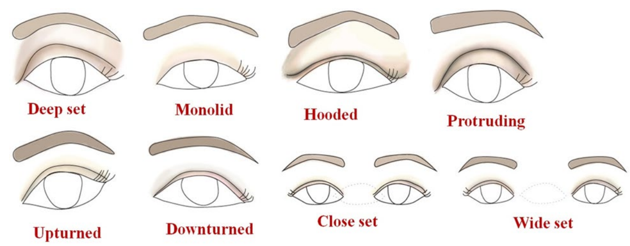


Fig. 4 Possible eye types [23]

3.1.4 Eyebrows type

There are nearly five types of category shown in Fig. 5. It is possible that each category consist of several sub categories. But for the sake of simplicity we are concerned about these main categories descriptions and representations provided below [24].

1. **Rounded** Rounded eyebrows are typically like a graph with respect to time which monotonically grows up to a level and slow down after some time. The end point of the graph is always above the height of the starting position.
2. **Hard angled** This represents the equal change in x and y which symbolize a line. This line grows up to a limit and falls in the same proportion. The width of the curve decreases from its slop from where the curve goes down.
3. **Soft angled** They are same like hard angled but the first half is more in width than the hard angled.
4. **"S" shaped** The width is almost identical to the soft angle but the shape represents a curve. This curve is having similarity with English alphabet "S".
5. **Flat** This is same like a flat curve with almost no changes. It gets affected only when we do any expression.

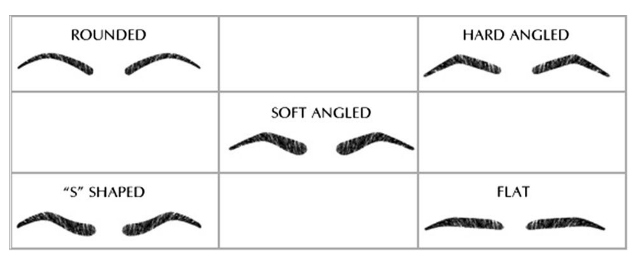


Fig. 5 Eyebrow types [24]

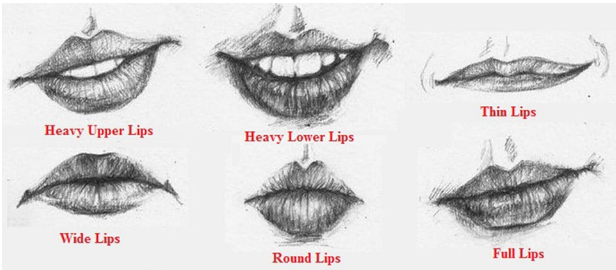


Fig. 6 Lip types [25]

3.1.5 Lip type

A complete list of lip types are presented in Fig. 6 [25]. There are total six categories of lips. These categories are formed based on the lip geometrical as well as physical property.

1. **Heavy upper lips and lower lip** The lower lip is dominated by upper lip and vice verse. It could be possible that they would be of varying size but the main factor would always be present.
2. **Thin lips** As the name suggests the width of both the upper and lower lips are very less. They can be easily classified.
3. **Full and wide lips** The by default type of the lips are full lips, if any property is dominated on this then the dominated category is assigned to the lip otherwise it is considered a full lips. If the lip width is more, then it will be treated as wide lips.
4. **Round lips** They are compact in size. They form a round shape.

3.1.6 Nose type

Different kinds of nose shapes are possible. The shape of the nose depends on climate. Their shape and size vary in order to adapt the climate change. The other reason responsible for the variability is due to the different size of ethmoid and maxillary sinuses in nose. These are the pockets below the eyes on both sides. In total we have explored nine types of nose, presented in Fig. 7 [24].

3.1.7 Special marks

Special marks can be of any cut on the face or mole or any birth mark. Sometimes these marks have special meaning based on the division of the face where they are located. In the realm of face reading the human face could be divided among 130 individual macro physical features or three micro features [26]. These micro features are named as celestial region (upper zone), self-will region (middle zone) and earthly region (lower zone). The shape and size of each zone are varying from person to person. Each zone has its unique property like upper zone explains about the imagination power of a person, middle zone states about the memory and lower zone describe about the observation quality of the person [26]. The fragmentation mechanism discussed in face reading realm motivated us to quantify the problem by dividing the face into four regions. Here we are not saying that face can only be divided into four regions. We have divided it to four regions to reduce the complexity as well as to keep the maximum discriminative power. These are recorded in the form of their location type of special mark and the size of mark. An example is shown in Fig. 8.

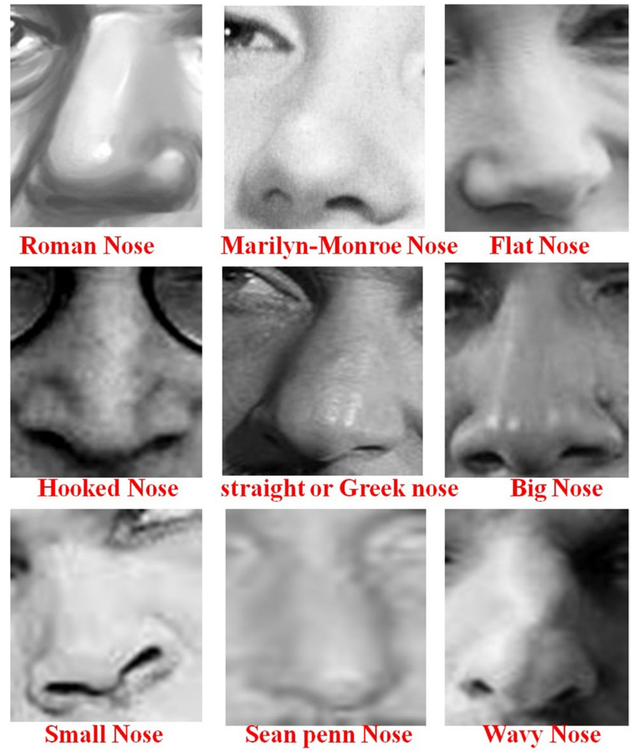


Fig. 7 Nose types [24]

3.2 Knowledge representation

All features and their values extracted in the previous unit of knowledge acquisition are raw and cannot be used directly in system modelling. Therefore we need to transform these values over the set of simplified (symbolic) values. These transformations help in setting the domain for each attribute set. Symbolic feature transformation for each feature set is discussed below.

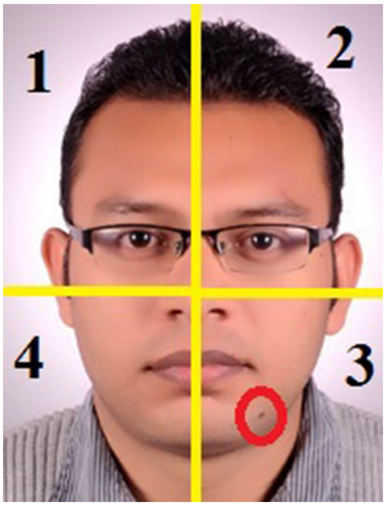


Fig. 8 Special marks

Table 1 Transformation for gender type

Attribute type	Transformed values
Male	M
Female	F
Domain size	1–2

3.2.1 Gender type

Gender is key attribute which can reduce our search space. If the criminal is male we will search for only male criminals and vice verse. The transformation is shown below in Table 1.

3.2.2 Face type

There are eight possible categories for face. Each category is symbolized by a numeric value shown in Table 2. In this case the domain for face attribute will be (1–8).

In a nutshell we have used eight types of attribute to define a person. One physical attribute and seven facial attribute. Each attribute has its own domain size and way of representation. But for making the whole system simple we have generally used a linear transformation [12]. Table 3 shows the summarized representation of these attribute. This table helps us to recall the attribute definition and domain.

4 Knowledge processing: a rough set based modeling

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 need to be processed by rough set theory(RST) [27–29]. RST is sufficient enough to handle the uncertainty and imprecision [30, 31]. There are four basic steps involved in rough set modelling [32–34] (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.

Table 2 Transformation for face type

Attribute type	Transformed values	Attribute type	Transformed values
Oblong	1	Square	5
Rectangle	2	Diamond	6
Round	3	Inverted triangle	7
Oval	4	Triangle	8
Domain size	1–8		

Table 3 Attributes, labels and domains: a summarized representation

Attribute name	Attribute symbol	Domain
Gender	a_1	1–2
Face type	a_2	1–8
Face tone	a_3	1–20
Eye shape	a_4	1–9
Eyebrows type	a_5	1–5
Lip type	a_6	1–6
Nose type	a_7	1–9
Special marks	a_8	1–13

4.1 Construction of information system

The information system (IS) is composed of the universe of discourse U and the attribute set A [27, 33, 34]. It is represented by:

$$IS = (U, A) \tag{1}$$

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 ∈ 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. In our model the information system consists of 300 classes and for each class five training views and two testing views exist. Therefore, we have 1500 objects and 8 attributes to describe them. These eight attributes

Table 4 An information system instance

Objects (U)	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	Decision attribute (d)
o_1	M	3	1	1	3	5	1	1	2
o_2	M	2	1	2	3	5	1	1	2
o_3	F	1	1	1	3	6	3	8	3
o_4	M	2	3	1	2	4	2	1	1
o_5	F	4	1	1	4	7	3	8	3
o_6	M	2	1	2	3	5	1	1	2
o_7	M	2	3	1	2	4	2	1	1
o_8	F	1	1	1	3	6	3	8	3
o_9	M	4	3	1	2	5	2	1	1
o_{10}	F	1	1	1	3	6	3	8	3

1500 samples from the criminal dataset, 300 classes

Table 5 An elementary set description for attribute set A

U/A	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	d
o_3, o_8, o_{10}	F	1	1	1	3	6	3	8	3
o_4, o_7	M	2	3	1	2	4	2	1	1
o_2, o_6	M	2	1	2	3	5	1	1	2
o_1	M	3	1	1	3	5	1	1	2
o_5	F	4	1	1	4	7	3	8	3
o_9	M	2	1	2	3	5	1	1	2

description is presented in Table 3. 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 Table 3. For demonstrating our hypothesis, we have randomly selected only ten objects to represent the IS as shown in Table 4.

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 $[o_i]_{IND_{(A)}}$. It may so happen that we are interested about only specific attribute set. Then the elementary set description will be different. Let $B \subset A$ consists of attribute set $\{a_1, a_2, a_3, a_9\}$. In this case the elementary set created on attribute set B will be different as shown in Tables 5 and 6, respectively.

4.2 Finding core and reducts

The attribute used to represent the system is further processed and analysed 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 [35, 36]. These attributes will later help to setup the rules library. The core can be found as the set of all singleton entries in the discernibility matrix. The discernibility matrix is a square matrix (n × n) where n is the number of elementary sets. Before proceeding for core and reducts, some key ingredients,which will be useful in this section shall be discussed.

Table 6 An elementary set description for attribute set B

U/A	a_1	a_7	a_8	d
o_3, o_5, o_8, o_{10}	F	3	8	3
o_2, o_6	M	2	1	2
o_4, o_7	M	1	1	1
o_1	M	1	1	2
o_9	M	2	1	1

1. **Lower, upper and boundary approximation** There could be three conditions arising for predicting the data over one of the elementary set which will be expressed by these three categories [31, 34]. *Lower approximation:* elements possibly belong to the elementary set. *Upper approximation:* elements those certainly belong to the elementary set. *Boundary approximation:* elements those belong to the boundary sets. Let O is the

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

$$\underline{BO} = \{o_i \in U | [o_i]_{IND(B)} \subset O\} \tag{2}$$

Similarly the upper approximation of B is defined as:

$$BO = \{o_i \in U | [o_i]_{IND(B)} \cap O \neq \phi\} \tag{3}$$

Any object which belongs to the lower set can also belong to the upper set.The boundary approximation is defined by the intersection of lower and upper approximation. It can be represented as:

$$BNO = BO - \underline{BO} \tag{4}$$

Example 1 Let O is a group of some objects O = $\{o_2, o_3, o_4, o_6, o_7\}$ and over the attribute set A (please refer to Table 11).
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:

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

2. **Approximation accuracy** Accuracy of these approximations is also represented by rough membership (μ) and it is calculated using:

$$\mu_B(O) = card(\underline{BO})/crad(BO) \tag{5}$$

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(BO) = 4 and card(BO) = 7 Then Accuracy: $\mu_B(O) = 4/7 = 0.57$

3. **Independence of attributes** The attributes used here for representing the information system have dependency on others or they may be independent. If the attribute is dependent, we can remove, therefore, minimize the attribute set. The dependent attributes can be discovered by removing it from the system and calculating the number of elementary sets. If the number of elementary sets are identical as before (in the absence of particular attribute). This attribute is called superfluous attribute. An example is shown below in Table 7. From Table 7 we can see that in first elementary set each attribute is superfluous with respect to the whole attribute set. So we can remove any of them. But when we remove the combination of two attributes as shown in second elementary set, only $a_1 a_2$ remain superfluous, rest become dispensable. Similarly by removing three attributes $a_1 a_2 a_3$ act as superfluous.
4. **Computing core and reducts** The concept of dispensable attribute gives rise to the concept of reducts. Reducts are the set of attributes which give out the same level of classification as well as preserve the partitions. Reducts (RED) are computed with the help of discernibility matrix. The discernible relation of two objects o_i and o_j shows how they discern from each other based on their attribute set. Core is the common attribute/attributes among the reduct sets. We represent core as:

$$CORE(B) = \cap RED(B) \tag{6}$$

An example for calculating the core and reducts is demonstrated in Table 8. Here the sets are same as we have already calculated over attribute set A (refer to Table 5) The set description we used in above table is as follows:

Table 7 Attribute dependency and elementary sets

	Removed attributes								
	None	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
Number of elementary Sets	6	6	6	6	6	6	6	6	6
	Removed attributes								
	None	a_1, a_2	a_2, a_4	a_2, a_6					
Number of elementary sets	6	6	5	5					
	Removed attributes								
	None		a_1, a_2, a_3	a_2, a_5, a_6	a_2, a_4, a_6				
Number of elementary sets	6		6	4	4				

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 discernibility function f(A) is applied to construct the reducts for the calculated discernible matrix. Here we use five discernibility functions for each elementary set. Like $f_1(A)$take care of only about how the elementary set 1 differs from 2, 3, 4, 5 and 6.

Table 9 Elementary set description over attribute set C

U/C	a_2	a_5	a_6	d
o_3, o_8, o_{10}	1	3	6	3
o_4, o_7	2	2	4	1
o_2, o_6	2	3	5	1
o_1	3	3	5	2
o_5	4	4	7	2
o_9	4	2	5	1

These are the sets of optimal attribute which result in the same partition. Tables 9 and 10 show an instance of this. (C = $a_2a_5a_6$ and $D = a_2a_5a_8$)

$$
\begin{aligned}
f_1(A) &= (a_1 + a_2 + a_3 + a_5 + a_6 + a_7 + a_8) \times (a_1 + a_2 + a_4 + a_6 + a_7 + a_8) \\
&\quad \times (a_1 + a_2 + a_6 + a_7 + a_8) \times (a_2 + a_5 + a_6) \times (a_1 + a_2 + a_3 + a_5 + a_6 + a_7 + a_8) \\
&= (a_2 + a_5 + a_6) \text{ (Corresponding to 1st row of Table 14)} \\
f_2(A) &= (a_3 + a_4 + a_5 + a_6 + a_7) \times (a_3 + a_2 + a_5 + a_6 + a_7) \times (a_1 + a_2 + a_3 + a_5 + a_6 + a_7 + a_8) \times (a_6) \\
&= (a_3 + a_5 + a_6 + a_7) \times (a_6 + a_8) \\
f_3(A) &= (a_2) \times (a_1 + a_2 + a_3 + a_4 + a_5 + a_6 + a_7 + a_8) \times (a_3 + a_4 + a_5 + a_2 + a_7) \\
&= (a_2) \times (a_2 + a_4 + a_5 + a_7) \\
&= a_2 \\
f_4(A) &= (a_1 + a_2 + a_5 + a_6 + a_7 + a_8) \times (a_3 + a_2 + a_5 + a_7) \\
&= (a_2 + a_5 + a_7) \\
f_5(A) &= b(a_1 + a_3 + a_5 + a_6 + a_7 + a_8)
\end{aligned}
$$

The core of the set is calculated by:

$$f(A) = f_1(A) \times f_2(A) \times f_3(A) \times f_4(A) \times f_5(A) \tag{7}$$

$$
\begin{aligned}
f(A) &= (a_2 + a_5 + a_6) \times (a_3 + a_5 + a_6 + a_7) \times (a_6 + a_8) \times a_2 \times (a_2 + a_5 + a_7) \times (a_1 + a_3 + a_5 + a_6 + a_7 + a_8) \\
&= a_2 \times (a_6 + a_8) \times (a_2 + a_5 + a_6) \times (a_2 + a_5 + a_7) \times (a_3 + a_5 + a_6 + a_7) \times (a_1 + a_3 + a_5 + a_6 + a_7 + a_8) \\
&= a_2 \times (a_8 + a_10) \times (a_4 + a_7 + a_8) \times (a_4 + a_7 + a_9) \times (a_5 + a_7 + a_8 + a_9) \\
&= a_2 \times (a_6 + a_8) \times (a_2 + a_5 + a_6) \times (a_5 + a_7) \\
&= a_2 \times (a_6 + a_8) \times a_7 \text{ OR } a_2 \times a_6 \times (a_5 + a_7) \\
&= a_2 \times (a_5a_6 + a_5a_8) \text{ OR } a_2a_6 \times (a_5 + a_7) \\
&= a_2a_5a_6 + a_2a_6a_8 \text{ OR } a_2a_5a_6 + a_2a_5a_7
\end{aligned}
$$

There are three different reducts we achieved from the above discernibility matrix $\{a_2a_5a_6, a_2a_5a_7, a_2a_5a_8\}$.

5. **Classification** Reducts help to find out the minimal attribute set which help to represent the whole system. These reducts also help in classification. As the rule library is based on reducts, so before proceeding to generate rules for particular reduct, it is required to

Table 8 Discernibility matrix

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

Table 10 Elementary set description over attribute set D

U/D	a_2	a_5	a_8	d
o_3, o_8, o_{10}	1	3	8	3
o_4, o_7	2	2	1	1
o_2, o_6	2	3	1	1
o_1	3	3	1	2
o_5	4	4	8	2
o_9	4	2	1	1

Table 11 Classification accuracy over $\{a_2a_5a_6\}$

Class number	Number of objects	Lower approximation	Upper approximation	Accuracy
1	4	4	4	1
2	3	3	3	1
3	3	3	3	1

Table 12 Classification accuracy after removal of individual attribute

Class number	Remove one attribute			
	None	a_2	a_5	a_6
1	1.0	1.0	1.0	1.0
2	1.0	0.4	1.0	1.0
3	1.0	0.25	1.0	1.0

check the accuracy on the given classes [37, 38]. Let us have three classes as defined earlier in this section. Class 1: $\{o_3, o_5, o_8, o_{10}\}$ Class 2: $\{o_2, o_6, o_1\}$ Class 3: $\{o_4, o_7, o_9\}$ and reduct is C = $\{o_2, o_5, o_6\}$ The classification accuracy of each class over the attribute set $a_2a_5a_6$ is presented in Table 11. Here all the classes show a confidence of 100% classification which means we can proceed with this reduct set. Further we evaluate the accuracy of these classes after removing each individual attribute from this given attribute set. There is only one case as shown in Table 12 where classification accuracy of class 2 and class 3 decreases. Apart from this, rest of the classes show the same confidence as they show in Table 11.

4.3 Construction of decision table

Decision rules are constructed based on the decision table. A decision table is based on two types of attributes [37–39]. First group consists of those objects that agree on certain attributes and their values and other groups include all those objects who discern from the first group on that particular attribute and values. Let's take an example to understand this. We want to create a decision table for criminal # 1. $\{o_4, o_7, o_9\}$ belongs to first group as they agree on decision attribute d, while second group consists of $\{o_1, o_2, o_3, o_5, o_6, o_8, o_{10}\}$. The decision table shown in Table 13 presents all attributes which discern from each other over attribute set $\{a_2a_5a_8\}$.

From Table 13 we can see that object o_4 differ from o_1 over the attribute $a_4a_7a_{10}$, we can also notice that is difference is due to the values of $a_2 = 2, a_5 = 2$ and $a_8 = 1$. The rules formed on this table can be expressed using conjunction operator (∨), disjunction operator (∧) and implication operator (⇒).

$(a^2{}_2 \vee a^2{}_5 \vee a^1{}_8) \wedge (a^2{}_5) \wedge (a^2{}_2 \vee a^2{}_5 \vee a^1{}_8) \wedge (a^2{}_2 \vee a^2{}_5 \vee a^1{}_8) \wedge (a^2{}_5) \wedge (a^2{}_2 \vee a^2{}_5 \vee a^1{}_8) \wedge (a^2{}_2 \vee a^2{}_5 \vee a^1{}_8)$... $(Rule - 1)$

$a^2{}_2 \vee a^2{}_5 \wedge a^2{}_7 \wedge a^2{}_2 \vee a^2{}_5 \vee a^1{}_8 \wedge a^2{}_2 \vee a^2{}_5 \vee a^1{}_8 \wedge a^2{}_5 \wedge a^2{}_2 \vee a^2{}_5 \vee a^1{}_8 \wedge a^2{}_2 \vee a^2{}_5 \vee a^1{}_8$...$Rule - 2$

$a^4{}_2 \vee a^2{}_5 \wedge a^4{}_2 \vee a^2{}_5 \wedge a^4{}_2 \vee a^2{}_5 \vee a^1{}_8 \wedge a^2{}_5 \vee a^1{}_8 \wedge a^4{}_2 \vee a^2{}_5 \wedge a^4{}_2 \vee a^2{}_5 \vee a^1{}_8 \wedge a^4{}_2 \vee a^2{}_5 \vee a^1{}_8$...$Rule - 3$

The problem with this rule library is that it consist lots of redundancy. This redundancy can be minimized using traditional Boolean algebra. The expression:

$(a^2{}_2 \vee a^2{}_5 \vee a^1{}_8) \wedge (a^2{}_5) \wedge (a^2{}_2 \vee a^2{}_5 \vee a^1{}_8) \wedge (a^2{}_2 \vee a^2{}_5 \vee a^1{}_8) \wedge (a^2{}_5) \wedge (a^2{}_2 \vee a^2{}_5 \vee a^1{}_8) \wedge (a^2{}_2 \vee a^2{}_5 \vee a^1{}_8)$... $(Rule - 1)$

Corresponds to object o_4 and o_7. This rule makes an implication that if (a) and (b) hold true then the criminal is identified as criminal # 1.

$$\begin{aligned}(a_2 = 2) \wedge (a_5 = 2) &\Rightarrow C_1\\ (a_8 = 1) \wedge (a_5 = 2) &\Rightarrow C_1\\ (a_5 = 2) &\Rightarrow C_1\end{aligned} \tag{8}$$

Similarly the expression

$a^4{}_2 \vee a^2{}_5 \wedge a^4{}_2 \vee a^2{}_5 \wedge a^4{}_2 \vee a^2{}_5 \vee a^1{}_8 \wedge a^2{}_5 \vee a^1{}_8 \wedge a^4{}_2 \vee a^2{}_5 \wedge a^4{}_2 \vee a^2{}_5 \vee a^1{}_8 \wedge a^4{}_2 \vee a^2{}_5 \vee a^1{}_8$

will be summarized as:

Table 13 Decision table

	o_1	o_2	o_3	o_5	o_6	o_8	o_{10}
o_4	$a^2{}_2, a^2{}_5, a^1{}_8$	$a^2{}_5$	$a^2{}_2, a^2{}_5, a^1{}_8$	$a^2{}_2, a^2{}_5, a^1{}_8$	$a^2{}_5$	$a^2{}_2, a^2{}_5, a^1{}_8$	$a^2{}_2, a^2{}_5, a^1{}_8$
o_7	$a^2{}_2, a^2{}_5$	$a^2{}_7$	$a^2{}_2, a^2{}_5, a^1{}_8$	$a^2{}_2, a^2{}_5, a^1{}_8$	$a^2{}_5$	$a^2{}_2, a^2{}_5, a^1{}_8$	$a^2{}_2, a^2{}_5, a^1{}_8$
o_9	$a^4{}_2, a^2{}_5$	$a^4{}_2, a^2{}_5$	$a^4{}_2, a^2{}_5, a^1{}_8$	$a^2{}_5, a^1{}_8$	$a^4{}_2, a^2{}_5,$	$a^4{}_2, a^2{}_5, a^1{}_8$	$a^4{}_2, a^2{}_5, a^1{}_8$

$= ({a^4}_2 \vee {a^2}_5) \wedge ({a^2}_5 \vee {a^1}_8) \wedge ({a^4}_2 \vee {a^2}_5 \vee {a^1}_8)$
$= ({a^4}_2 \vee {a^2}_5) \wedge \{({a^2}_5 \vee {a^1}_8) \wedge ({a^4}_2 \vee {a^2}_5 \vee {a^1}_8)\}$
$= ({a^4}_2 \vee {a^2}_5) \wedge \{({a^2}_5 \vee \{({a^1}_8) \wedge ({a^4}_2 \vee {a^1}_8)\}\}$
$= ({a^4}_2 \vee {a^2}_5) \wedge \{({a^2}_5 \vee \{({a^1}_8 \wedge {a^4}_2) \vee ({a^1}_8)\}\}$
$= ({a^4}_2 \vee {a^2}_5) \wedge \{{a^2}_5 \vee ({a^1}_8 \wedge {a^4}_2)\}$
$= ({a^4}_2 \vee {a^2}_5) \wedge \{({a^2}_5 \vee {a^1}_8) \wedge ({a^2}_5 \vee {a^4}_2)\}$
$= ({a^4}_2 \vee {a^2}_5) \wedge ({a^2}_5 \vee {a^1}_8)$

The following rule is obtained from this deduced expression:

$$\begin{aligned} &(a_2 = 4) \wedge (a_8 = 1) \Rightarrow C_1 \\ &\qquad (a_5 = 2) \Rightarrow C_1 \end{aligned} \tag{9}$$

From expression 8 and 9 we conclude that the criminal # 1 is identified as:

$$\begin{aligned} &(a_2 = 2) \wedge (a_5 = 2) \Rightarrow C_1 \\ &(a_8 = 1) \wedge (a_5 = 2) \Rightarrow C_1 \\ &(a_2 = 4) \wedge (a_8 = 1) \Rightarrow C_1 \\ &\qquad (a_5 = 2) \Rightarrow C_1 \end{aligned} \tag{10}$$

These rules can be further explained as:

- if face is RECTANGLE and eyebrow is HARDANGLED then criminal is # 1.
- if face is OVAL and face cut mark is NOMARK then criminal is # 1.
- if eyebrow is HARDANGLED and face cutmark is NOMARK then criminal is # 1.
- if eyebrow is HARDANGLED then criminal is # 1.

4.4 Rules verification/validation

There are four basic parameters suggested by [38] and [39] to check the validity of each rule. These four parameters are: (a) support of the rule (b) certainty factor (c) strength of rules and (d) coverage.

1. **Support** Generally the rule is defined as $C \Rightarrow_o D$. Here C is the set of conditional attribute and D is the decision attribute. If we define a rule say $(a_2 = 2) \wedge (a_5 = 2) \Rightarrow C_1$ then the support of this rule includes all objects (o) which belong to the universe of discourse U, that follow this rule. The expression for calculating support of a rule is expressed as: $SUPP_o(C, D) = |C(o) \cap D(o)|$. Example: The support for each rule is presented in Table 14.
2. **Strength** Strength of the rule is calculated based on total objects and support of the rule. It is expressed as: $\sigma_o(C, D) = (\frac{SUPP_o(C,D)}{(|U|)}).|U|$ is the cardinality of the objects. The example is shown in Table 14.
3. **Certainty** The certainty factor $cer_o(C, D)$ is defined as the ratio between support of the rule and $|C|$. $|C|$ represents the number of objects grouped over the conditional attribute (C). Certainty factor is calculated as follows: $cer_o(C, D) = \frac{(|C(o) \cap D(o)|)}{(|C(o)|)} = \frac{(SUPP_o(C,D))}{(|C(o)|)}$. Example is presented in Table 14.
4. **Coverage** Coverage of the rule is computed over the decision attribute D. It is the ratio between the support of the rule and the Decision attribute D. If a class consists of n samples/objects. Then the coverage of the rule is considered as the ratio between support and sample per class. Expression used in calculation is as follows: $cov_o(C, D) = \frac{(|C(o) \cap D(o)|)}{(|D(o)|)} = \frac{(SUPP_o(C,D))}{(|D(o)|)}$. Example is presented in Table 14.

Table 14 Classification accuracy after removal of individual attribute

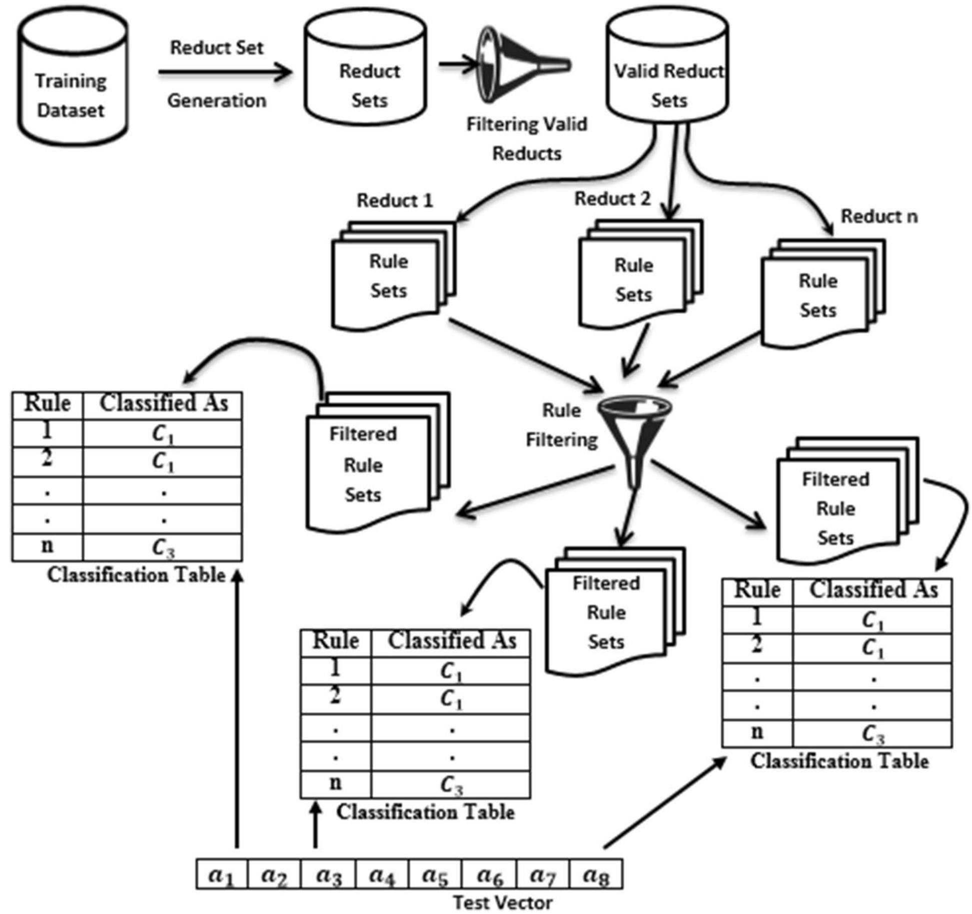
Rules	Support	Strength	Certainty	Coverage
1	2	0.2	1.0	0.67
2	3	0.3	1.0	1.0
3	1	0.1	1.0	0.33
4	3	0.3	1.0	1.0

Score are computed based on Table 11

5 Experimental setup

The experimental setup consists of several modules. The layout of our experiment is presented in Fig. 9. The training module is involved to generate different reduct and rule sets. These reducts sets are further filtered based on their classification accuracy. Rule library for each reduct is set in the next module. As we can see the rule library has been defined for each individual reduct. We have lot of rules; therefore a filter is also placed to remove the unwanted rules. Certainty and coverage are the two factors used to evaluate each individual rule. Only those rules which have coverage and certainty factor greater than the threshold qualify as valid rules. These rules are saved to be used later for projecting the criminal identity. The second part of this module is testing. A test vector is supplied to the entire rule library. Based on the attribute values different rules from different rule library classify the test vector to different classes. For each rule library we keep the status of which rule classified the test vector to which class in the form of a classification table. Each library has its own classification table. These rule libraries are further analysed as

Fig. 9 Experimental setup for RST based criminal identification



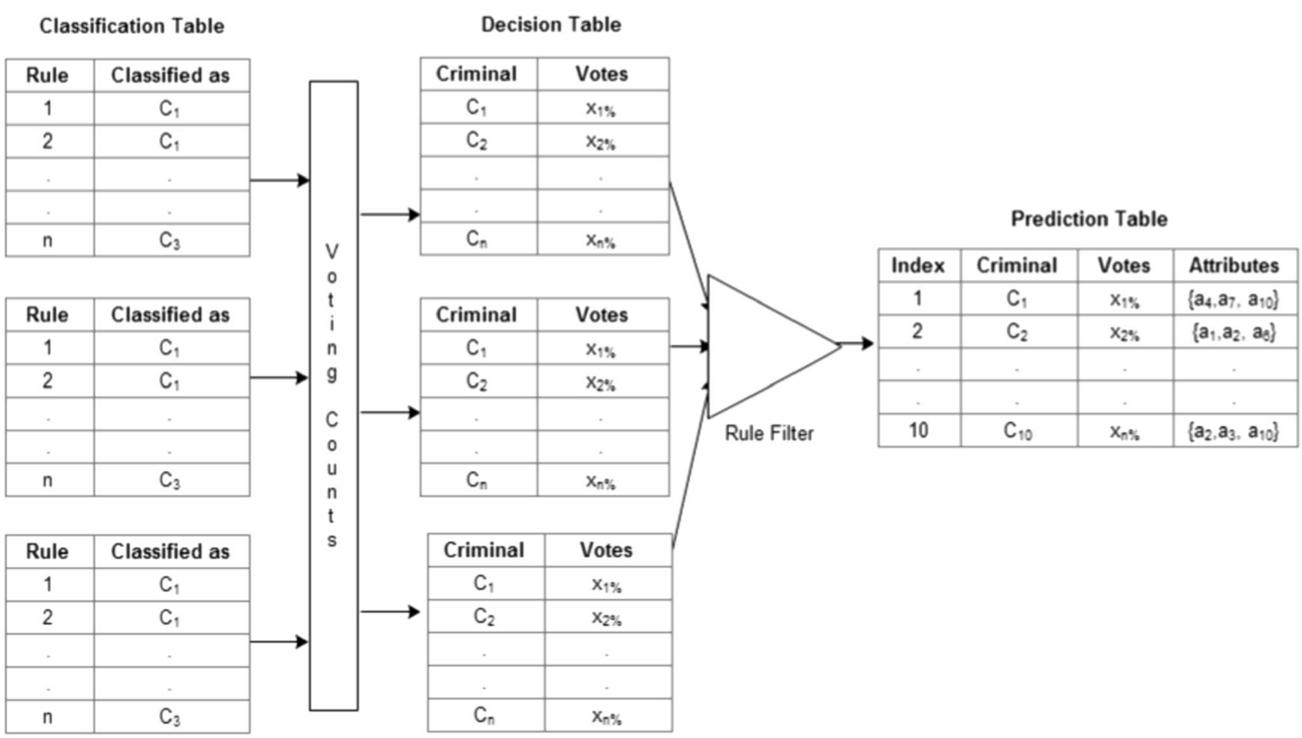


Fig. 10 Experimental setup for RST based criminal identification

shown in Fig. 10 to get the actual level of the test vector. The actual class label is based on the most occurring label from all classification tables. Let us assume that we have 3 rule libraries L1, L2 and L3 and 3 classification tables T1, T2 and T3 respectively. Also assume that criminal # 1 is the majority voted name from T1 and T3, while T2 shows the most occurred name as criminal # 2. In this case the final class label is given as criminal # 1, as two classification tables support this. Moreover a prediction table shown in Fig. 10 is created which consists of top ten nearest matches based on all classification tables. The rank is assigned based on the votes given to the particular criminal in the respective classification table. Let's take a scenario when classification table 1 has 75 rules. Out of 75 rules 50 rules classified the test identity to criminal # 1. This means decision table 1 has 50/75 = 66.37% votes in favour of criminal # 1. Similarly votes from all the classification table is generated and analysed to find top ten votes and concerned identities. Later these results are summarized in a prediction table shown in Fig. 10. The rest modules of this experiment are briefly discussed in subsections.

5.1 Database generation

There are three different datasets that we use to demonstrate the effectiveness of rough set based modelling, over mug-shot detection. The glimpse of the dataset is given in Fig. 11.We have randomly selected 50 colour photographs of 50 females and 50 males from FERET dataset. The RGB images are taken because there we can define the face skin tone as one of the attribute of the information system. We select only one image per person. A similar kind of strategy we followed to select face images from the AR database while we use all 100 photo faces from the CUHK dataset. In total we have 300 criminal faces in our database. This work is totally based on human perception. Therefore, we have invited several views from different volunteers to create our dataset. Views are collected with the help of our designed web interface shown in Fig. 1. The interface displays the photograph of the guilty and captures the user response through input fields. Prototype of each facial attribute (like what kind of face type you have or eye shape) is added to assist the volunteers views. We have represented each criminal to one class; therefore we have 300 classes and per class 5 training samples. Similarly we have 2 test samples per class in testing database. In total we have used 105 volunteers to mark the criminals identity in a symbolic way. Each volunteer is given the description of about 20 criminals. We created 15 groups, each group has 7 members. Hence for each group we have 140 views. Each criminal has 5 training views and test views. The training dataset is filled up with 1500 objects (criminal descriptions) each one can be expressed with the help of 8 attributes. Similarly the test set consists of 600 views of the same population. We have used different volunteer views for training and testing. The volunteers used here to generate training and testing views are the students of the Indian Institute of Information Technology Allahabad, India.

5.2 Reduct sets and rule generation

The reducts set generated on 1500 objects are presented in Table 15. There exist 300 classes (unique decision attributes) and each class consist of 5 samples.There are 78 reduct sets. Most of the reduct sets are having average attribute value of 3 moreover, they don't have any core. Each reduct set is later filtered out to remove some unwanted reducts. We keep only those reducts which have a classification accuracy of 100%. After applying the filtering process we end up with only 32 reducts. The reduct set helps in generating the rule library. The number of rules attached with each rule set is presented in Table 15. The numbers of rule are huge. Many rules are redundant and need simplification. The simplified version of these rules is shown in Table 16. Each rule is tested against certainty and coverage factor. If the certainty and coverage factor are greater than 90%, we consider the rules.

5.3 Classification of test vector

There are 600 test views and two test views per class. The test vector is supplied to each filtered rule library. We used MATLAB parallel processing toolbox to process the test vector over different rule sets in parallel. Although each rule in the rule library classifies the test vector to one of the classes, we see which class has maximum number of rule support. The ratio between the number of rules and the number of rules supported to a particular class is known as votes (%). Whichever class has the highest vote (%) will be considered as the output of the particular rule library. The final result is summarized in Table 17 in the form of top ten results. These results are based on hierarchy of maximum votes. Example is shown in Fig. 12.

6 Results and discussion

The results are analysed for how many test cases were taken and how many of them classified correctly. We have tested the existing system over two different conditions. In first condition we show a face from the training database and ask person to give her/his views about the person. So, in this scenario we have liberty to see and decide the

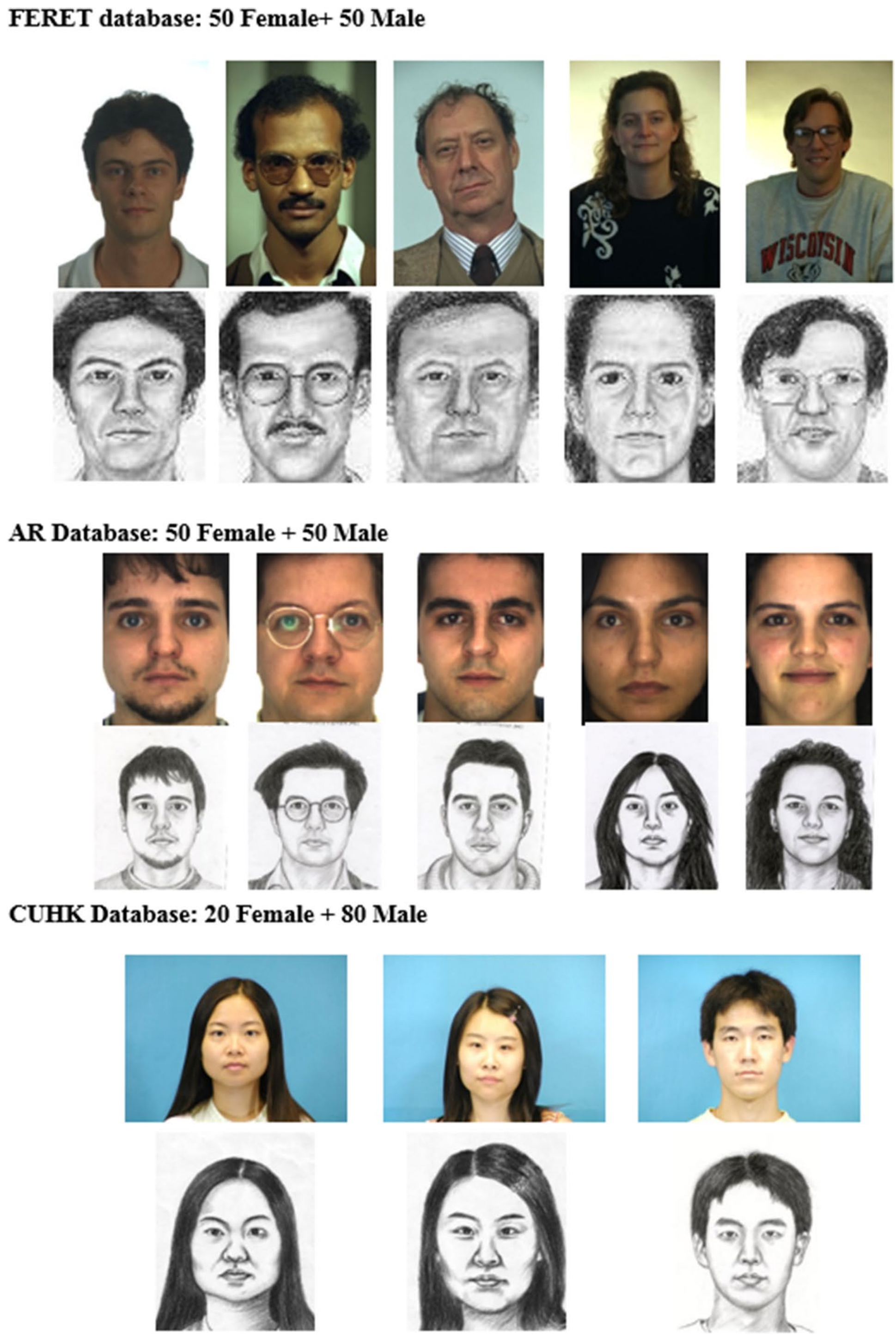


Fig. 11 Mug-shot database

presented criminal attribute like face type, eye shape etc. In second test condition we made these photos disappear from the interface as shown in Fig. 12. We kept an unknown identity photograph and placed a caption of that identity (please refer Fig. 13). So, they have to make a guess about captioned person.

In both the cases we reported different classification results. In test case one, we achieved 92.67% accuracy. This shows that out of 600 test cases 556 test cases are classified correctly. In other word we can say that the predicted criminal comes into top 10 results. The results are summarized in Figs. 14 and 15. Figure 14 depicted how many test

Table 15 Reduct set with their accuracy and respective number of rules

Reducts	Rules	Classification accuracy (%)
(a_2, a_5, a_6)	725	100
(a_1, a_3, a_7)	685	100
(a_1, a_5, a_8)	615	100
(a_1, a_3, a_5)	580	100
(a_1, a_4, a_5)	470	100
(a_1, a_2, a_5, a_8)	410	100

Table 16 Reduced rule sets for each reduct

Reducts	Filtered rules
(a_2, a_5, a_6)	489
(a_1, a_3, a_7)	457
(a_1, a_5, a_8)	433
(a_1, a_3, a_5)	418
(a_1, a_4, a_5)	395
(a_1, a_2, a_5, a_8)	380

cases are misclassified. The classification has the value one and the misclassification has value 0. Bins are representing the misclassified points. In our second case the accuracy decreases and reached up to 73.83%. The results are summarized in Fig. 15. Moreover in both the cases the top 1 result which is rated as highest voted criminal turned to be true with the accuracy as 63.68 and 59.21%.

6.1 Perception VS mug-shot detection

A comparative study between the mug-shot detection approaches and the proposed perception based criminal identification is presented in Table 17. An analysis is taken separately for each individual database. The results are summarized in Table 17 for both the test cases. We achieve encouraging classification accuracy in test case-I and test case-II.

6.2 Discussion

We had two test cases in our experiment. In the first case the photograph of the criminal was just in front of the person while in the second case he/she saw that photograph just 1 day before. Based on these test cases we reported different results. We figure out the reason behind these differences and found that this is due to the different level of perception maturity [40, 41]. In first case when the criminal's photograph is directly visible to us, our perception is stronger than the case where it is absent. This clearly shows that our accuracy is fully dependent on user's views (perception) [42, 43]. The stronger is the perception the more accurate is the result. Further, we were interested in knowing how human mind stores the information about any stimulus. How long it is stored in the brain and at what rate it vanishes from the brain. Our brain has three different sections for perception, attention and storage of information. These sections and their connectivity are depicted in Fig. 16. Special sensory interneuronal connections are involved in case of perception capturing unit while Thalamus, Frontal lobe and prefrontal cortex are responsible for

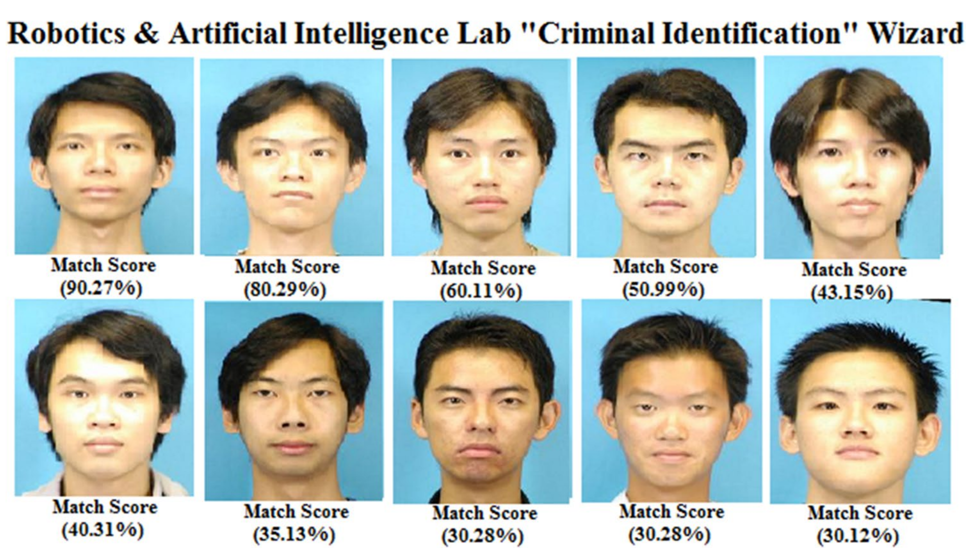


Fig. 12 Top 10 results of criminals

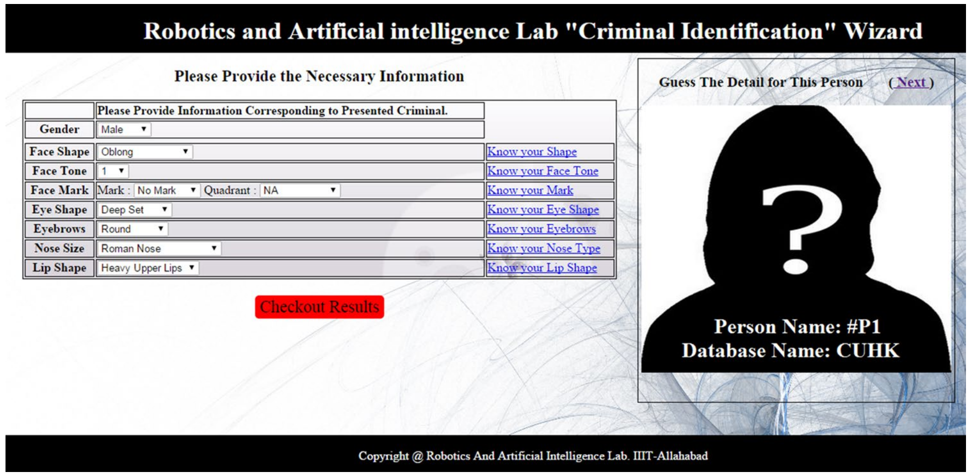


Fig. 13 Interface without criminal face (test case-II)

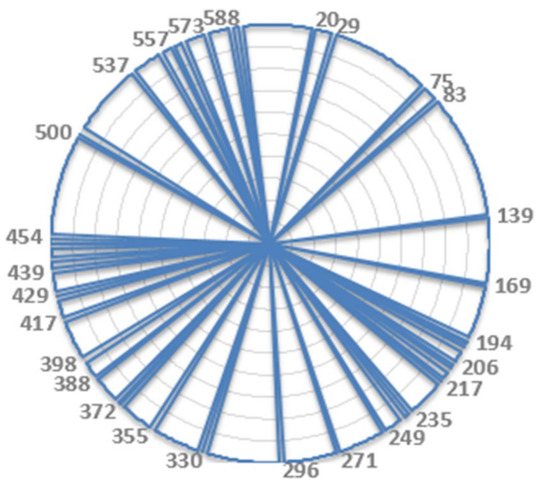


Fig. 14 Misclassification on test case-I

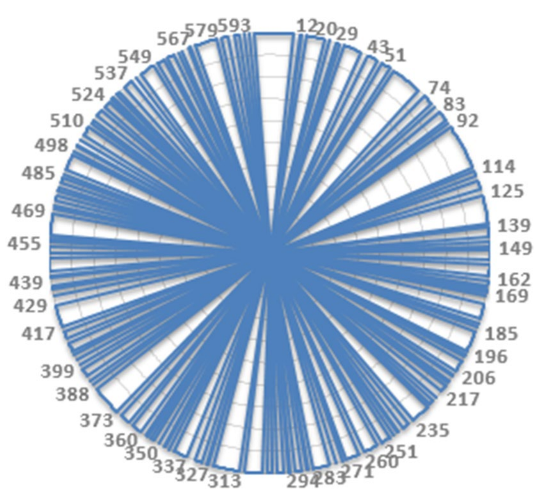


Fig. 15 Misclassification on test case-II

attention. Hippocampus and Medial temporal lobe stores information about stimuli.

Memory is the ability to encode, store and recall an event. Processes involved in human memory are therefore encoding, storage and recall or retrieval [40]. There is a fourth memory consolidation which could be a part of encoding process or storage process or it could be a separate entity in its own right as memory stabilization after initial acquisition. Perceived sensations are decoded in various sensory areas of cortex through neural network connection. There is no one single organ for memory processes. Moreover excessive emotional arousal leads to attention deficit for other details. Therefore, at the crime scene subject (eyewitness) notes the weapon but criminals face and dress and vehicle pass unnoticed. There are also three different kinds of memory we have in our brain listed in Table 18 [40]. Their snapshot can be seen in Fig. 16. Table 18 clearly shows that if the eyewitness had the criminal face or physical information in his/her long term memory, it will give better results.

7 Conclusion

This paper demonstrates a new technique to identify criminals in the early stage of investigation through Eyewitness Visual Perception. It has been noted that this perception is imprecise, uncertain and vague in most of the cases. In order to tackle these challenges, a rough set theory based modelling is performed. We generated 32 reducts to pursue this experiment. Classification accuracy is considered as the metric to select valid results. Several performance metrics such as strength, confidence and coverage are also estimated to deduce the rule library associated with each reduct. The strength of this approach

Table 17 Comparatives analysis with existing approaches (test case-I and II)

Authors	Population	Test case-I (%)	Test case-II (%)
Tang et al. [14]	188 sketch and photo pair (Train: 88, Test: 100)	96	
Liu et al. [16]	606 sketch and photo pair (306 Train + 300 Test)	97	
Gao et al. [13]	606 sketch and photo pair (306 Train + 300 Test)	98.89	
Wang et al. [10]	606 sketch and photo pair (306 Train + 300 Test)	99.70	
Proposed approach	500 Training Views + 200 Testing Views (same dataset as used in CUHK [10])	94.50	87.50
Proposed approach	500 Training Views + 200 Testing Views (same dataset as used in AR[10])	96	79
Proposed approach	500 Training Views + 200 Testing Views (same dataset as used in FERET[10])	89	73

Table 18 Memory type and store duration

Memory type	Memory duration
Ultra short memory	200–500 ms
Short term memory	10–15 s to 1 min
Long term memory	Life long

is evaluated in the light of 2100 samples (1500 training + 600 testing) of our domestic symbolic database. The domestic database is created on the basis of 300 pseudo criminals (CUHK dataset). Each criminal is treated as one class and for each class exists 5 training and 2 testing views. We considered two different cases to measure the efficiency of our proposed model. In the first case criminal photograph is visible at the time of knowledge capturing while in the second phase we have shown the same photograph 1 day before. In both cases the accuracy is reported around 92.67 and 73.83% respectively. The difference in accuracy could be due to the different level of perception maturity of the viewer. Different level of maturity is due to the different memory functions of the brain. If the criminal face is stored in eyewitness long term memory, then this will give better result in comparison to short term memory.

The future work includes the use of additional attributes like age, height, weight, hair style, eyeball colour etc. to produce better matches. As a future scope this module could be applied prior to using mug-shot detection module. The strength of this proposed work is that only eyewitness notions are required to identify the criminal which avoids the complexity of making the sketch.

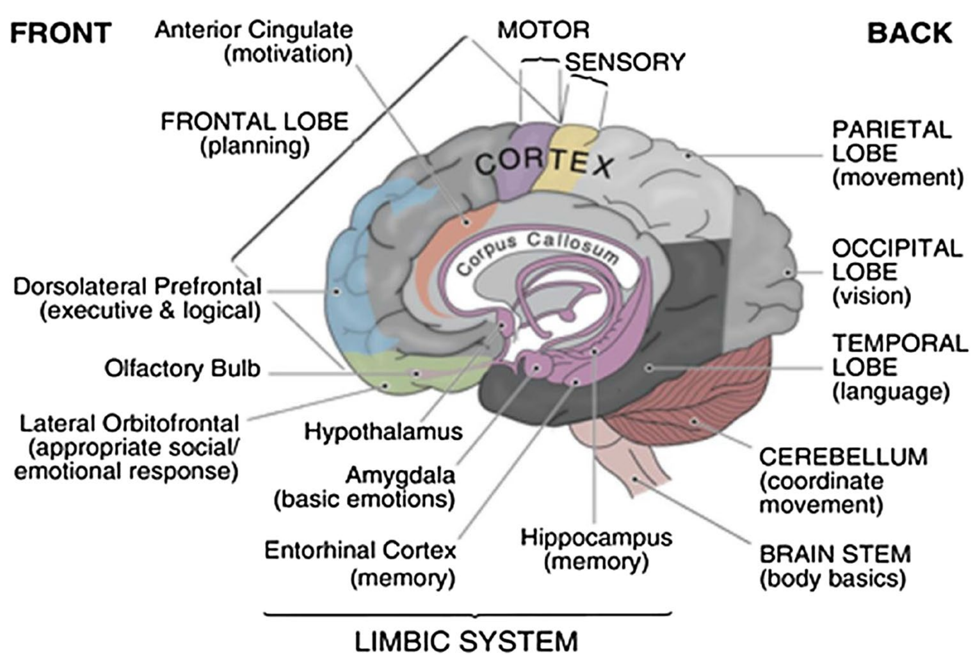


Fig. 16 Brain memory structure and interconnection [40]

Acknowledgements Author would like to thank Dr V. Joshi former doctor at the Banaras Hindu University (Varanasi), India for explaining the memory processes and neuro anatomy of limbic system. Gratitude is given to all the students of Robotics and Artificial Intelligence Lab Indian Institute of Information Technology Allahabad, India and those enrolled for HUR class. Special thanks to Chinese University of Hong Kong for providing their mug-shot database to accomplish this study.

References

1. Jain AK, Klare B, Park U (2012) Face matching and retrieval in forensics applications. MultiMedia IEEE 19(1):20
2. Jain AK, Klare B, Park U (2011) Face recognition: some challenges in forensics. Automatic face and gesture recognition and workshops (FG 2011), 2011 IEEE international conference on, pp 726–733
3. Gibson L (2008) Forensic art essentials. Elsevier, Oxford
4. Taylor K (2001) Forensic art and illustration. CRC Press, Boca Raton
5. Koshimizu H, Tominaga M, Fujiwara T, Murakami K (1999) On KANSEI facial image processing for computerized facial caricaturing system PICASSO. Systems, man, and cybernetics, 1999. IEEE SMC '99 conference proceedings. 1999 IEEE international conference on, vol 6, pp 294–299
6. Chen H, Xu Y-Q, Shum H-Y, Zhu S-C, Zheng N-N (2001) Example-based facial sketch generation with non-parametric sampling. Computer vision, 2001. ICCV 2001. In: Proceedings of eighth IEEE international conference on, vol 2, pp 433–438
7. Wang J, Bao H, Zhou W, Peng Q, Yingqing X (2002) Automatic image-based pencil sketch rendering. J Comput Sci Technol 17(3):347–355
8. Uhl RG, Lobo NDV (1996) A framework for recognizing a facial image from a police sketch. In: Proceedings of IEEE international conference, computer vision and pattern recognition
9. Wang X, Tang X (2009) Face photo–sketch synthesis and recognition. Pattern Anal Mach Intell IEEE Trans 31(11):1955–1967
10. Klare B, Li Z, Jain AK (2011) Matching forensic sketches to mug-shot photos. IEEE Trans Pattern Anal Mach Intell 33(3):639–646
11. Xiao B, Gao X, Tao D, Li X (2009) A new approach for face recognition by sketches in photos. Signal Process 89(8):1576–1588
12. Singh AK, Nandi GC (2016) Visual perception-based criminal identification: a query-based approach. J Exp Theor Artif Intell 29(1):175–196
13. Gao X, Zhong J, Tao D, Li X (2008) Local face sketch synthesis learning. Neurocomputing 71(1012):1921–1930
14. Tang X, Wang X (2003) Face sketch synthesis and recognition. In: Proceedings of IEEE internatiaonl conference, computer vision
15. Tang X, Wang X (2004) Face sketch recognition. IEEE Trans Circuits Syst Video Technol 14(1):50–57
16. Liu Q, Tang X, Jin H, Lu H, Ma S (2005) A nonlinear approach for face sketch synthesis and recognition. In: Proceedings of IEEE conference on computer vision and pattern recognition, San Diego, CA, USA, pp 1005–1010
17. Xiao B, Gao X, Tao D, Yuan Y, Li J (2010) Photo-sketch synthesis and recognition based on subspace learning. Neurocomputing 73(46):840–852
18. Gudivada VN, Raghavan VV, Seetharaman GS (1993) An approach to interactive retrieval in face image databases based on semantic attributes. In: Proceedings of third annual symposium document analysis information retrieval
19. Klare BF, Klum S, Klontz JC, Taborsky E, Akgul T, Jain AK (2014) Suspect identification based on descriptive facial attributes. In: Biometrics (IJCB), 2014 IEEE international joint conference on IEEE, pp 1–8
20. Siddiquie B, Feris RS, Davis LS (2011) Image ranking and retrieval based on multi-attribute queries. In: Computer vision and pattern recognition (CVPR), 2011 IEEE conference on IEEE, pp 801–808
21. Mane DR, Kale AD, Bhai MB, Hallikerimath S (2010) Anthropometric and anthroposcopic analysis of different shapes of faces in group of Indian population: A pilot study. J Forensic Legal Med 17(8):421–425
22. http://en.wikipedia.org/wiki/VonLuschanschromaticscale. Accessed 29 Sept 2014
23. http://www.beautylish.com/a/vcacj/whats-your-eye-shape. Accessed 29 Sept 2014
24. http://design.tutsplus.com/tutorials/human-anatomy-fundamentals-advanced-facial-features-cms-20683. Accessed 29 Sept 2014
25. Gudivada VN, Raghavan VV, Seetharaman GS (1993) An approach to interactive retrieval in face image databases based on semantic attributes. In: Proceedings of third annual symposium of document analysis information retrieval
26. Kuei CA (1998) Face reading. M. Evans and Company, New York
27. Pawlak Z (1982) Rough sets. Int J Parallel Program 11(5):341–356
28. Pawlak Z (2002) Rough sets and intelligent data analysis. Inf Sci 147(14):1–12
29. Leung Y, Wei-Zhi W, Zhang W-X (2006) Knowledge acquisition in incomplete information systems: a rough set approach. Eur J Oper Res 168(1):164–180
30. Dntsch I, Gediga G (1998) Uncertainty measures of rough set prediction. Artif Intell 106(1):109–137
31. Kryszkiewicz M (1998) Rough set approach to incomplete information systems. Inf Sci 112(14):39–49
32. Komorowski J, Pawlak Z, Polkowski L, Skowron A (1998) Rough sets: a tutorial. In: Rough fuzzy hybridization: a new trend in decision-making, pp 3–98
33. Pawlak Z, Skowron A (2007) Rudiments of rough sets. Inf Sci 177(1):3–27
34. Pawlak Z (2004) Elementary rough set granules: toward a rough set processor. Rough-neural computing. Springer, Berlin
35. Swiniarski RW (2001) Rough sets methods in feature reduction and classification. Int J Appl Math Comput Sci 11(3):565–582
36. Khoo L-P, Tor S-B, Zhai L-Y (1999) A rough-set based approach for classification and rule induction. Int J Adv Manuf Tech 15:438–444
37. Qian Y, Liang J, Dang C (2008) Converse approximation and rule extraction from decision tables in rough set theory. Comput Math Appl 55(8):1754–1765
38. Pawlak Z (1997) Rough set approach to knowledge-based decision support. Eur J Oper Res 99(1):48–57
39. Qian Y, Liang J, Li D, Zhang H, Dang Chuangyin (2008) Measures for evaluating the decision performance of a decision table in rough set theory. Inf Sci 178(1):181–202
40. Koriat A, Goldsmith M, Pansky A (2000) Toward a psychology of memory accuracy. Annu Rev Psychol 51(1):481–537
41. Eisen ML, Morgan DY, Mickes L (2002) Individual differences in eyewitness memory and suggestibility: examining relations between acquiescence, dissociation and resistance to misleading information. Personal Indiv Differ 33(4):553–571
42. Sarwar F, Allwood CM, Innes-Ker Å (2014) Effects of different types of forensic information on eyewitness memory and confidence accuracy. Eur J Psychol Appl Legal Context 6(1):17–27
43. Forgas JP, Laham SM, Vargas PT (2005) Mood effects on eyewitness memory: affective influences on susceptibility to misinformation. J Exp Soc Psychol 41(6):574–588