Journal of Experimental & Theoretical Artificial Intelligence

ISSN: 0952-813X (Print) 1362-3079 (Online) Journal homepage: https://www.tandfonline.com/loi/teta20

Visual perception-based criminal identification: a query-based approach

Avinash Kumar Singh & G. C. Nandi

To cite this article: Avinash Kumar Singh & G. C. Nandi (2017) Visual perception-based criminal identification: a query-based approach, Journal of Experimental & Theoretical Artificial Intelligence, 29:1, 175-196, DOI: 10.1080/0952813X.2015.1132266

To link to this article: https://doi.org/10.1080/0952813X.2015.1132266

Published online: 07 Feb 2016.

JOURNAL OF EXPERIMENTAL & THEORETICAL ARTIFICIAL INTELLIGENCE, 2017
VOL. 29, NO. 1, 175–196
http://dx.doi.org/10.1080/0952813X.2015.1132266

Taylor & Francis
Taylor & Francis Group

Visual perception-based criminal identification: a query-based approach

Avinash Kumar Singh and G. C. Nandi

Robotics & Artificial Intelligence Laboratory, Indian Institute of Information Technology, Allahabad, India

ABSTRACT

The visual perception of eyewitness plays a vital role in criminal identification scenario. It helps law enforcement authorities in searching particular criminal from their previous record. It has been reported that searching a criminal record manually requires too much time to get the accurate result. We have proposed a query-based approach which minimises the computational cost along with the reduction of search space. A symbolic database has been created to perform a stringent analysis on 150 public (Bollywood celebrities and Indian cricketers) and 90 local faces (our data-set). An expert knowledge has been captured to encapsulate every criminal's anatomical and facial attributes in the form of symbolic representation. A fast query-based searching strategy has been implemented using dynamic decision tree data structure which allows four levels of decomposition to fetch respective criminal records. Two types of case studies - viewed and forensic sketches have been considered to evaluate the strength of our proposed approach. We have derived 1200 views of the entire population by taking into consideration 80 participants as eyewitness. The system demonstrates an accuracy level of 98.6% for test case I and 97.8% for test case II. It has also been reported that experimental results reduce the search space up to 30 most relevant records.

ARTICLE HISTORY
Received 26 November 2014
Accepted 25 November 2015

KEYWORD
Criminal identification; dynamic query optimisation; knowledge extraction; knowledge representation; classification; reasoning; expert system

1. Introduction

A person can be categorised from their behavioural, physiological and facial attributes or the combination of all three. A person's behaviour can be captured in terms of their gait patterns and the way of speaking, their physic can be represented in the form of age, gender and height, while the face description can be summarised as face category, face tone, eyebrows type, eye shape, nose shape and lip size. This paper utilizes all these attributes and talks about its contribution over the forensic domain more precisely on criminal identification. The working envelop of forensic consists of two major blocks: evidence collection and secondly the correct matching of these evidences (Gibson, 2010; Jain, Klare, & Park, 2012). Face can be treated as one of the kernel evidences; therefore its proper matching is required. But it is a known true fact that if someone is committing a crime then she or he will never want to disclose her/his identity. In such scenarios, we would be having only partial information of the faces or may be no face information at all. A detailed list of these challenges was addressed by Jain, Klare and Park (2011). Due to parallel processing of human brain and its logical capability, it is easier for us to recognise any person with the given partial information. But this is not true in case of machines. Please

CONTACT Avinash Kumar Singh rs110@iiita.ac.in; G. C. Nandi gcnandi@iiita.ac.in

refer (Chellappa, Wilson, & Sirohey, 1995) to know more about how human perception and machine react towards face recognition problem and their synergism too.

We started our journey of criminal identification following the work of Wells (1978). According to Wells, investigations and prosecutions of criminal cases heavily depend on the eyewitness testimony. Eyewitness testimony plays a supportive role in order to achieve criminal justice. Generally an eyewitness has to follow a lineup to identify criminal in the bunch of suspects. In such scenarios, those who are not culprit are known as fillers. But, sometime the impreciseness and vague perception about the criminal could raise conditions like (a) target absent and (b) target present. Target absent is when the culprit is added into fillers and target present is when culprit is excluded away from fillers. The whole process of identification could be processed following two different methodologies. Simultaneous lineup is where all suspicious members are present in front of the eyewitness, while sequential lineup is when suspects are shown one by one. The literature also categorises the eyewitnesses into two categories, genuine and mock witness. Genuine witnesses are those who have seen the event live while mock witnesses are people who have actually not seen the event or were not present at the site based on the eyewitnesses' verbal description; they are asked to pick a person in a lineup. There are other factors which can influence the identification accuracy and can affect the criminal justice. Factors which can affect the justice process are known as system variables while others are known as estimator variables. The system variables are the collection of set of protocols such as the instruction given prior to viewing the lineup, lineup size etc. while the estimator variable includes the evaluation of current witnessing situation and the relation between the culprit and the eyewitness. In most of the cases the identification is performed by seeing the face and other physiological as well as behavioural characteristics of the suspect. Therefore, in this paper we have quantified face and physiological characteristics to identify criminals. As face falls under the system variable category, it is an integral part of jurisdiction. It is widely useful in such scenario where we don't have any idea about the criminal. Face recognition is a tool which could be helpful to trace out the victim by crawling the existing criminal database and to find the match. But it is applicable only when we have photograph of the criminal. When we don't have photograph of the criminal, an eyewitness who has seen the guilty can help us. Usually it helps sketch artist to prepare the sketch of the criminal. These sketches are the embroidery of eyewitness imagination and perception on the piece of paper. In most of the cases, these sketches are similar to the criminal face. Later these sketch images are published publicly by the police department to find the guilty. But the problems of manual matching of sketches to photographs are ambiguous and need automation.

The existing literature defines automatic way of matching these sketches to their respective photo images as mugshot detection. The first step towards mug shot detection started in 1996 with Uhl and da Vitoria Lobo (1996). In his experiment, they demonstrated the sketch to photo matching successfully on a limited database. Later the significant contribution in this direction has been carried out by Wang and Tang (2009), Xiao, Gao, Tao, and Li (2009), and Klare, Li, and Jain (2011).Unfortunately, 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. The existing literature is also not utilising the other two features (behavioural and physiological) to trace out the victim. The addition of these two additional attribute can improve the recognition accuracy. 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 representation of imprecise knowledge of eyewitness. The same can also be recorded by verbal communication with eyewitness. This paper presents a study done on a semi-automated system requiring expert guidance to predict possible matches of criminal photos based on eyewitness imprecise knowledge. There are two major modules in the proposed system shown in Figure 1. The first module (training module) labels each criminal face with the help of some symbolic transformations. The criminal face database is labelled with the help of a beauty expert. The database is labelled on the basis of criminal's facial features such as face shape, face colour, face tone, eyebrow type, eye type, nose type, lip tone, lip type and anthropomorphic features such as gender, height, age. At first these facial attributes are filled up by the beauty expert after analysing criminal photos while the anthropomorphic attribute values are already known (generally

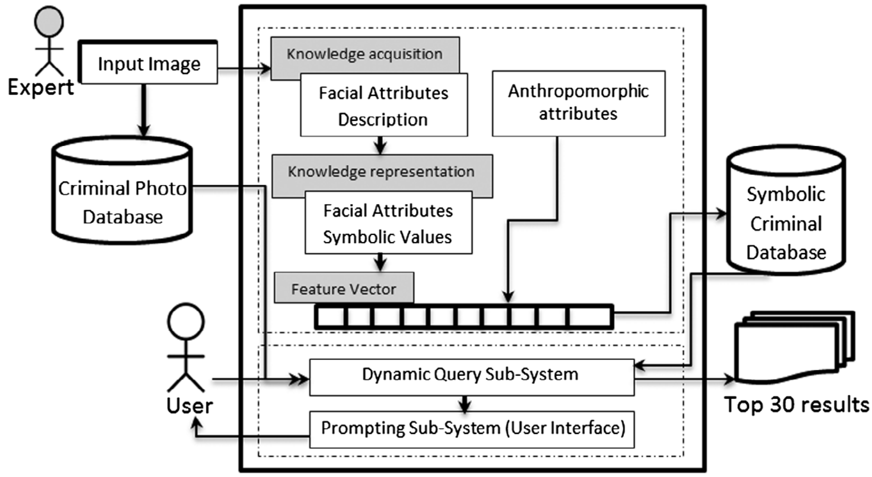

Figure 1. Proposed system architecture.

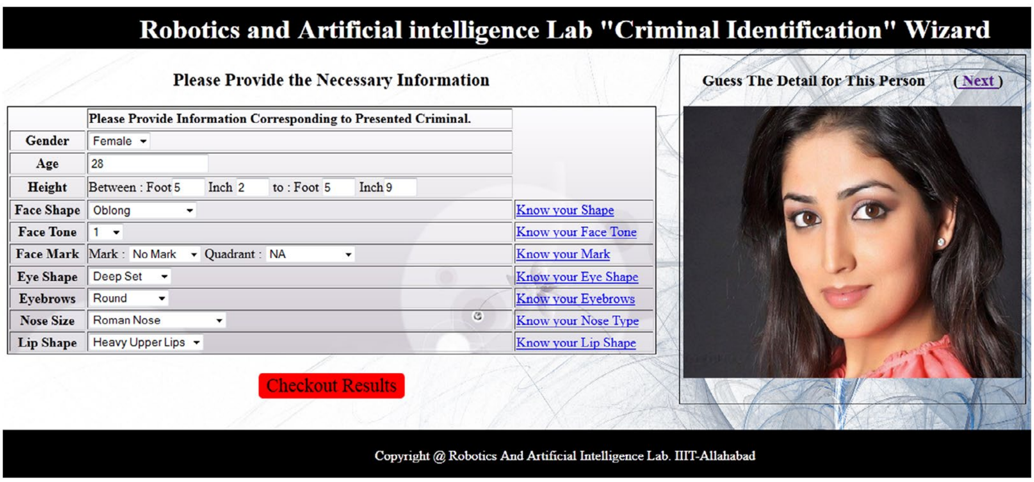

Figure 2. Interface of the proposed system.

the anthropomorphic attributes are measured by the police when criminal is arrested before). These values are further transformed by the knowledge representation module. These transformed values are known as the feature vector.

For each criminal face, one feature vector is stored in the symbolic criminal database. Both the criminal photo database and the criminal symbolic database are connected with the help of reference key. The second module (testing module) also captures the imprecise knowledge of the eyewitness with the help of our designed web interface shown in Figure 2. The interface asks different queries about suspect physical appearance and facial attribute. We believe that a person can be categorised and differentiated based on these attributes (Klare et al., 2014; Siddiquie, Feris, & Davis, 2011). Each attribute has different types associated with it which is required to represent any subject. Eyewitness can select the appropriate attribute type. The collected information is modelled using a query-based system to find the best possible matches. As a result, the system shows top 10, top 20 and top 30 results from the photo database. A testing framework is designed which consists of 240 so-called criminal faces which includes 110 Indian Bollywood celebrities data-set, 40 Indian Cricket team data-set and 90 from Robotics and Artificial Intelligence Lab data-set. This module is discussed in detail in Section 2.

The rest of the paper is summarised as follows: Section 2 describes analysis of previous research, Section 3 presents the design issues of the system followed by Section 4 which summarises the experimental modelling of query-based system. In the end, Section 5 presents proof of our hypothesis in terms of results and discussion. Finally, Section 6 concludes the paper with advantages, limitation and future scope of this new way of criminal identification.

2. Analysis of previous research

The first step towards sketch to photo matching was presented by Uhl and da Vitoria Lobo (1996). Although the experiment is performed on little database of 7 sketches and 17 real photographs, they managed to prove that this process can be further enhanced. Later, the significant contribution in this direction has been carried out by Gao, Zhong, Tao, and Li (2008), Xiao et al. (2009), Wang and Tang (2009). Most of the researchers (Tang & Wang, 2003; Uhl & da Vitoria Lobo, 1996; Xiao et al. 2009) used Karhunen–Loeve Transform to transform photo images to eigensketches. These eighensketches are pretty much similar to the sketch images. Thereby, they are minimising the gap between the sketch and the photo images. It is an accepted truth that whenever a sketch artist design any sketch, it is obvious that there would be the presence of distortions such as blur and overshading effect, thus causing ambiguity in the facial attributes. If I_p represents the photo image and I_s represents the sketch image, then the mapping which will transform the photo image to sketch image could be defined as: $I'_s = T(I_p)$. The aforementioned facts act as inhibitors and increase the gap between I'_s and I_s. In all those scenarios, the mapping T has to be non-linear in order to provide the better transformation such that $I'_s \cong I_s$. Hence the solution proposed by Uhl and da Vitoria Lobo, 1996, Tang and Wang, 2002, and Xiao et al. 2009 is not very much suitable in these circumstances. This motivated Liu, Tang, Jin, Lu, and Ma (2005) to work further in this direction and he introduced the concept of local geometry preserving non-linear 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 non-linear discriminant analysis. The same problem of non-linearity is also solved by Embedded Hidden Markov Model (E-HMM)-based Facial Sketch Synthesis by Gao et al. (2008). They used the same philosophy of Wang to generate the pseudo eigensketches but using different HMMs. They assumed that the transformation from photo images to pseudo sketches is linear hence the linear transformation was required. This non-linear transformation is modelled with E-HMM. Later the modified version of this approach has been seen in the work of Xiao et al. (2009) and Xiao, Gao, Tao, Yuan, and Li (2010). Although these approaches addressed the problem of non-linearity effectively, 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. (2010) by introducing the notion of subspace learning to generate synthesised photo sketches.

The previous literature is attacking the problem of sketch to actual photo prediction based on the eigenspace transformation but recently in 2011 Klare et al. (2011) enlightened the new ways of recognition. They used the fundamental techniques 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 by 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 addresses the first type of issue. Both of these cases are addressed by Klare et al. (2011). They tested the validity of their proposed approach and figure 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 representation of imprecise knowledge of eyewitness. The same can also be recorded by non-verbal communication with eyewitness.

The other category of literature deals with the descriptive facial attributes for criminal identification (Gudivada, Raghavan, & Seetharaman, 1993; Klare et al., 2014; Siddiquie et al., 2011). Recently, Klare et al. (2014) 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 expressed 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 3 category values (nose size small, normal and large), 6 attributes have 4 category values and other 2 attributes have 6 and 7 category values. The category value for each attribute is used to quantify that attribute. The efficiency of the proposed system has been demonstrated based on the Chinese University of Hong Kong face sketch data-set. The data-set consists of 1196 face sketches which are grouped over the gallery and probe images based on the twofold cross validation. Both gallery and probe images are labelled with respect to the defined 46 attributes by the persons as well the computer-designed algorithm. The accuracy of only attribute-based face recognition is not showing good results when tested individually while when it is combined with the previous sketch-based face recognition system (sketchId), it improves the accuracy from 84 to 92%. The other approach analyses semantic knowledge of face images to retrieve culprit photograph from the face database (Gudivada et al., 1993). 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 attribute's 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. Objects surrounding the person which influences the decision-making of the person are called elements or entities. However, the properties of the objects which affect the decision-making are known as construct or cognitive dimension. A rating is assigned to each construct which defines the degree of the construct. There are three values defined such as (a) value 1 for construct is certainly present, (b) value 2 for subject is neutral position and (c) value 3 if the construct is absent. There are 19 attributes defined to represent each face image and each attribute is defined in terms of the aforementioned construct rating. The relation between attribute assets is stored in the repertory grid. The repertory grid stored dependency values between two attributes. Higher the dependency values, higher the chances of getting included in the query. Other researchers have used multi attribute for image ranking and image retrieval (Siddiquie et al. 2011). They demonstrate the automatic extraction and labelling of different facial, geographical and anatomical features of a person. There are 27 features used to represent the person information. These features include the origin (nationality), face colour, hair colour, eyeglasses, 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 of this project. In order to extract valuable knowledge from the user, we have created a web interface. The web interface asks various queries about the criminal physical and facial attributes. A snap shot of the designed system is shown here in Figure 2.

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 (generally 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. Based on these attributes type, they can be categorised into four different sets (Mane, Kale, Bhai, & Hallikerimath, 2010). (a) Visual features: set of features like eye, nose, facial outline. (b) Anthropometric landmarks: special points and location of the human face which is considered by doctors for their study and experiments. (c) Cephalometric landmarks: shape and size of skull. (d) Special features: they are the special marks or cut on the face which usually does not match with other person's face marks. We have tried to extract all four category of feature set through our web interface. The feature description is presented one by one here.

A. *Face type:* Based on face anthropometric structure (Klare et al., 2014; Siddiquie et al., 2011) a face can be categorised over one of the classes shown in Figure 3, based on its outer boundary.
B. *Face tone:* Face tone speaks several things about your origin, your geographical origin etc. (Luschan, 2014). It can be described by a face tone parameter presented below in Figure 4.
C. *Eye shape:* As like the face tone and shape type, eye has also different kinds of shape, described below in Figure 5(Eye Shape, 2014).
   Deep set eyes: Deep set eyes are usually bigger in size and set a little deeper in the eye socket more salient brow bone.
   Monolid: Monolid kind of eyes represents flat with respect to facial anatomy. No stronger crease and not very much prominent brow bones visibility.
   Hooded eyes: Brow is covered with extra layer of skin and resultant upper eye lids look smaller.

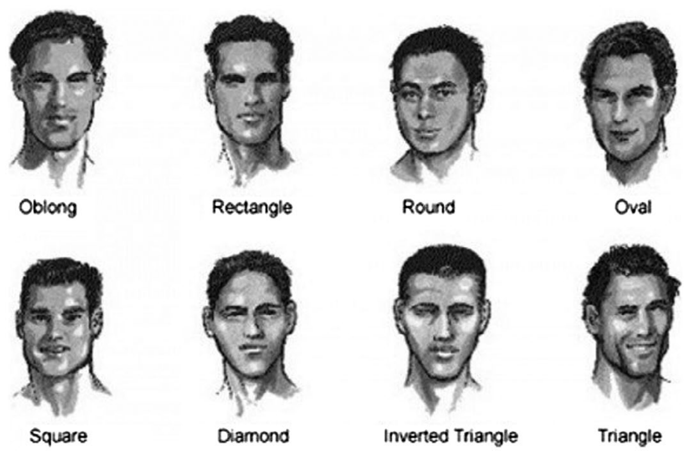

Figure 3. Different shapes of face (Human Anatomy, 2014).

	1	10			19	28	
	2	11			20	29	
	3	12			21	30	
	4	13			22	31	
	5	14			23	32	
	6	15			24	33	
	7	16			25	34	
	8	17			26	35	
	9	18			27	36	

Figure 4. Face skin tone (Luschan, 2014).

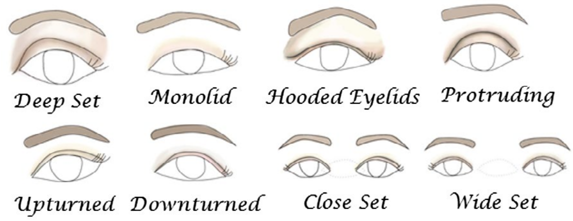

Figure 5. Possible eye types (Eye Shape, 2014).

Protruding eyes: Eye lids seem to be bulging from the eye socket. Such eye lids make eye attractive.
Upturned eyes: They look like an almond shape eye with the kick in outer corner.
Downturned eyes: They are same in shape as of upturned eyes only a drop in outer corner.
Close set eyes: When the distance between two eyes is less between the eye balls.
Wide set eyes: Just opposite the close set eyes, here the distance is more between eye balls.

D. *Eyebrows type:* There are nearly five types of category shown in Figure 6. It is possible that each category consists of several sub categories. But for the sake of simplicity we are concerned about these main categories' descriptions and representations provided below (Human Anatomy, 2014).
Rounded: Rounded eyebrows are typically like a graph with respect to time which monotonically grows up to a level and slows down after some time. The end point of the graph is always above the height of the starting position.
Hard angled: This represents the equal change in *x* and *y* which symbolise a line. This line grows up to a limit and fall in the same proportion. The width of the curve decreases from its slope from where the curve goes down.
Soft angled: They are same like hard angled but the first half is more in width than the hard angled.
"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".
Flat: This is same like a flat curve with almost no changes. It gets affected only when we do any expression.

E. *Lips type:* A complete list of lip types are presented in Figure 7 (Gudivada et al., 1993). There are total six categories of lips.
These categories are formed based on the lip geometry as well as physical property.
Heavy upper lips & lower lip: The lower lip is dominated by upper lip and vice versa. It could be possible that they would be of varying size but the main factor would always be present.

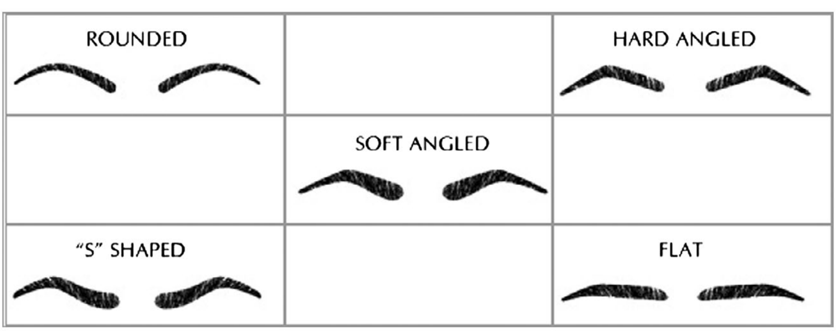

Figure 6. Eyebrow types (Human Anatomy, 2014).

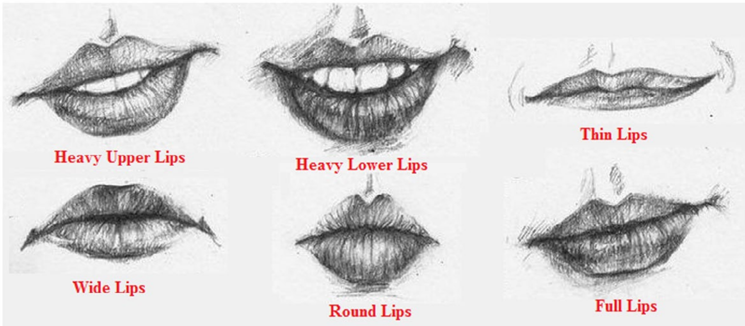

Figure 7. Lip shapes (Gudivada et al., 1993).

Thin lips: As the name suggests, the width of both the upper and lower lips are very less. They can be easily classified.
Full & wide lips: The full lips are the by default type, if any property is dominated on this then the dominated category is assigned to the lip otherwise it is considered a full lip. If the lip width is more then it will be treated as wide lips.
Round lips: They are compact in size. They form a round shape.

F. 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 Figure 8 (Human Anatomy, 2014).
G. *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 (Kuei, 1998). These micro features are named as celestial region (upper zone), self-will region (middle zone) and earthly region (lower zone).

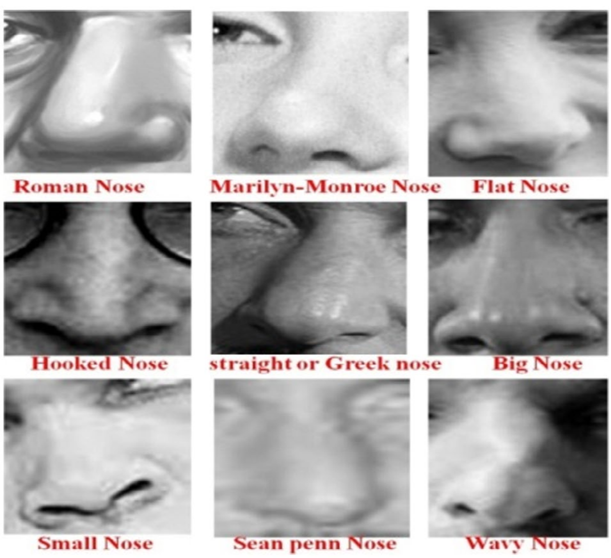

Figure 8. Nose types (Human Anatomy, 2014).

Figure 9. Special marks.

The shape and size of each zone varies from person to person. Each zone has its unique property like upper zone explains about the imaginative power of a person, middle zone states about the memory and lower zone describes the observation quality of the person (Kuei, 1998). The fragmentation mechanism discussed in face reading realm motivated us to quantify the problem by dividing the face over four regions. Here we are not claiming that face can only be divided over four regions. We have divided it over four regions to deduce 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 Figure 9.

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 transformation help in setting the domain for each attribute set. Symbolic feature transformation for each feature set is discussed below.

A. *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 versa. The transformation is shown below in Table 1.
B. *Face type:* There are eight possible categories for face. Each category is symbolised by a numeric value shown in Table 2. In this case the domain for face attribute will be (1–8).
C. *Face tone:* A wide variety of face tone is possible. It is not logical to assign particular values for each type. Therefore, with the help of a beauty expert we have categorised the face tone into several classes based on Von Luschan's chromatic scale (Luschan, 2014) shown in Table 3.
D. *Eye shape, nose type, lip type and eyebrow type:* As like the face type we can apply the linear transformation to Eye shape, Nose type, Lip type and Eyebrow type to decode Transformed values. The transformation is described in Table 4–7, respectively.
E. *Special marks:* Special marks also play an important role in identification. We have used three kinds of special marks (Birthmark, Mole and Cut mark) for the sake of representation. They can be placed anywhere in the face that's why we have divided the face into four quadrants. The quadrant is represented by Symbol 1/2/3/4 and marks are represented by Mole (M), Cut (C) and Birthmark (B), respectively. The transformation designed is summarised in Table 8.

In a nutshell, we have used 10 types of attributes to define a person: three physical attributes and seven facial attributes. 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. Table 9 shows the summarised representation of these attributes. This table helps us to recall the attribute definition and domain.

Table 1. Transformation for gender type.

Attribute type	Transformed value
Male	M
Female	F
Domain size [1–2]	

Table 2. Transformation for face type.

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

Table 3. Transformation for face tone using Von Luschan's chromatic scale.

Attribute type	Probable class	Transformed value
1–5	Very light or white	1
6–10	Light	2
11–15	Light intermediate	3
16–21	Dark intermediate or olive skin	4
22–28	Dark or brown type	5
29–36	Very dark or black type	6
Domain size [1–6]		

Table 4. Eye shape transformation.

Attribute type	Transformed value
Deep Set	1
Monolid	2
Hooded	3
Protruding	4
Upturned	5
Downturned	6
Close set	7
Wide set	8
Domain size [1–9]	

Table 5. Nose type transformation.

Attribute type	Transformed value
Roman nose	1
Monroe nose	2
Flat nose	3
Hooked nose	4
Greek nose	5
Big nose	6
Small nose	7
Sean penn nose	8
Wavy nose	9
Domain size [1–9]	

4. Experimental set-up

4.1. Database creation and labelling

This is the kernel step of our proposed experimental set-up. The database generation is divided into two modules. The first module is for collecting the public and local faces and their anatomical structure specification. We have used 110 Bollywood celebrities including 53 males and 57 females + 40

Table 6. Lip transformation.

Attribute type	Transformed value
Heavy upper lips	1
Heavy lower lips	2
Thin lips	3
Wide lips	4
Round lips	5
Full lips	6
Domain size [1–6]	

Table 7. Eyebrow transformation.

Attribute type	Transformed value
Round	1
Hard angled	2
Soft angled	3
"S" shaped	4
Flat	5
Domain size [1–5]	

Table 8. Special mark transformation.

Attribute type	Transformed value	Attribute type	Transformed value
No mark	1	3M	8
1M	2	3C	9
1C	3	3B	10
1B	4	4M	11
2M	5	4C	12
2C	6	4B	13
2B	7		
Domain size [1–13]			

Table 9. Attributes, labels and domains: a summarised representation.

Features	Attribute	Domain
f_1	Gender	[1–2]
f_2	Face type	[1–8]
f_3	Face tone	[1–6]
f_4	Eye shape	[1–9]
f_5	Eyebrow type	[1–5]
f_6	Lip type	[1–6]
f_7	Nose type	[1–9]
f_8.	Spcial mark	[1–13]
Class: 240 classes each class represent one criminal		

Indian cricketers + 90 students of Indian Institute of Information Technology, Allahabad to represent the database. Glimpse of the database is shown in Figure 10. Only those photographs are selected from the internet sources which have clear visibility for all the facial macro features. As we can see, the first row of Figure 10 represents the female population of Bollywood celebrities, the second row shows the male community. Cricketers and our in-home database are getting reflected in third and fourth rows, respectively.

The second module is to label the designed database using symbolic notations discussed in previous section. The guidance of the expert and some student volunteers has been used for database labelling. For example, given a picture of a lady shown in Figure 1, the expert can make a guess about her facial and anatomical attributes shown below.

Figure 10. Collection of public and local figure (The Pseudo Criminal Database).

Attributes	Gender	Age (years)	Height (cm)	Face type	Face tone	Eye shape	Eye-brows	Lip type	Nose type	Special mark
Values	Female	28	162	Oval	7	Hooded	Round	Full	Roman	No
Symbols	F	28	162	4	2	2	1	6	1	1

We also invited students of Indian institute of information Technology, Allahabad to contribute their views for the same. We are treating these volunteers as eye witnesses. Later the performance will be tested on the views collected from these eyewitnesses. A description about the labelled database is presented in Table 10.

The first group separated by a line in Table 10 symbolises Bollywood celebrities. The second group pointed to Indian cricketer's database and the third group shows records from our in house database.

4.2. Query design and optimisation

When the criminal database increases in size, it is required that only concerned records should be displayed to the user. Therefore, the query for extracting these records should be designed in such a way that it should give the minimal matched records (Jarke & Koch, 1984). It has been analysed and seen during several time testing on this project that all 10 features (please see Table-10 for feature specification) are not having the same weightage. Some special features have to be set which optimises the query and search results (Magerman, 1995). Here we have considered gender, age and height of the suspect as key features. Most of the records are filtered out by gender discrimination. Rest of these are further optimised by placing the age and height constraint. The fetched results depend on the window of age and height. More accurate the guess, more accurate search results. The search query will work as shown in Figure 11. The lower seven features $f_2, f_3, f_4, f_5, f_6, f_7, f_8$ are considered as the vague features and therefore we have assigned them to lower significance. The matching score is calculated based on

Table 10. Snapshot of the labelled database.

Subjects	Sex	Age (years)	Height (cm)	Face type	Face tone	Face mark	Eye type	Eye brows	Nose type	Lip shape
S1	F	33	162.56	1	3	1	4	3	5	2
S2	F	28	170.18	6	4	1	3	1	5	2
S3	F	26	175.26	4	3	1	4	2	9	4
S4	F	46	162.56	4	3	1	4	2	5	4
S5	F	21	165.10	1	3	1	1	4	9	4
S6	M	47	172.72	4	4	1	2	4	2	3
S7	M	32	172.72	6	3	1	2	1	2	2
S8	M	56	175.26	4	4	1	4	5	8	2
S9	M	70	199.64	4	4	1	6	4	5	4
S10	M	31	182.88	4	3	1	4	4	5	2
S11	M	26	175.00	3	3	1	3	2	3	3
S12	M	27	188.00	5	4	1	1	3	8	2
S13	M	41	165.00	4	4	1	1	4	4	1
S14	M	41	180.00	4	3	1	3	1	1	4
S15	M	27	182.88	6	4	1	2	4	2	3
S16	M	22	160.20	8	4	1	1	5	2	5
S17	M	22	180.34	2	4	1	2	3	8	3
S18	F	21	177.80	4	5	8	8	6	5	6
S19	F	23	165.10	5	3	3	5	4	2	4
S20	F	26	157.48	5	5	1	7	3	4	6

the hamming distance (Hamming Distance, 2014) between the train and test features. Two symbolic databases have been created to show the efficiency of the proposed system. The first database which is also known as the training database is established with the help of a beauty expert, while the second database is the collection of views of the general users. The query is crafted in such a way so that only the concerned record will come from the training database. Usually, the training database is crawled based on the test views about the criminal's gender, age and height attributes. The rest of the features such as face type (f_2), face tone (f_3), face mark (f_4), eye type (f_5), eye brows (f_2), nose type (f_2) and lip type (f_2) are projected as a result of these three pioneer features. The query used for extracting only concerned records is presented below.

$$\pi_{f_2,f_3,f_4,f_5,f_6,f_7,f_8}(\sigma_{\text{gender}=t_{\text{gender}} \text{ and age} \geq t_{\text{age}-5} \text{ \&\& age} \leq t_{\text{age}+5} \text{ and height} \geq t_{\text{height}-5} \text{ \&\& heigh} \leq t_{\text{height}+5}}(\text{DB}))$$

Here t_{gender}, t_{age}, t_{height} are the gender, age and height along the test views. We have created window of 5 years and 5 cm for age and height, because most of the time the approximation about the criminal's age and height is not accurate for test views. The above query will extract "*n*" number of records with respect to the given test features.

4.3. Matching technique

The matching or similarity between the train and test perception about the criminal is performed with the help of hamming distance. The hamming distance is calculated between the presented test sample and extracted "*n*" records from the training database based on the above query. If the test feature vector is symbolise by $\{ft_2, ft_3, ft_4, ft_5, ft_5, ft_6, ft_7, ft_8\}$ and the train features are represented by $\{f_2, f_3, f_4, f_5, f_6, f_7, f_8\}^i$ where *i* reflect the record number and it belongs to $i \in [1, n]$. Let's take an example how hamming distance will help to find a match. Whoever has the highest matching will be considered as the probable match case.

$$\text{FTest} = \{ft_2, ft_3, ft_4, ft_5, ft_5, ft_6, ft_7, ft_8\}$$

For instance let the test features set have the values.

$$\text{FTest} = \{1, 3, 1, 4, 3, 5, 2\} \tag{1}$$

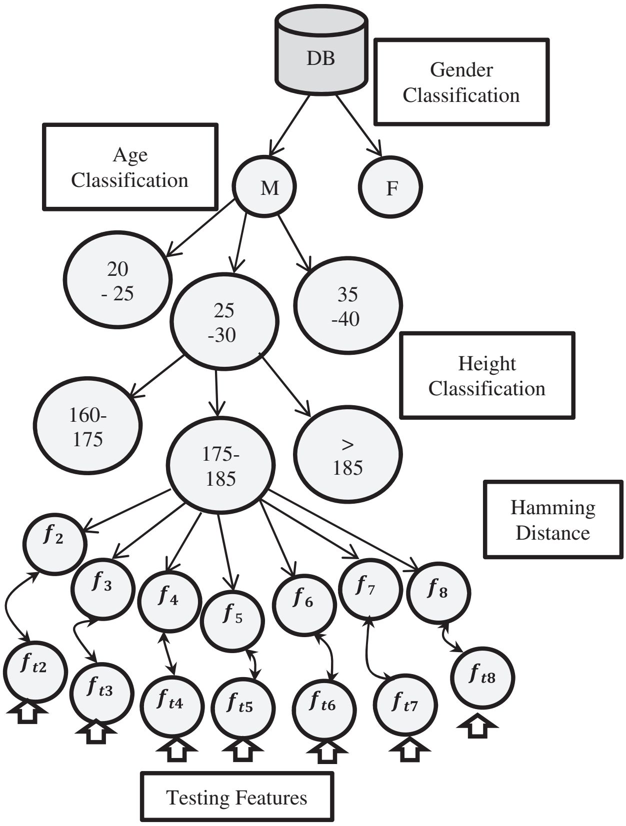

Figure 11. An instance of query optimisation.

Similarly training set have:

$$\text{FTrain} = \{f_2, f_3, f_4, f_5, f_6, f_7, f_8\}$$

For instance let the train features set have the values.

$$\text{FTrain} = \{1, 3, 1, 1, 4, 9, 4\} \tag{2}$$

The hamming distance between (1) and (2) is basically the number of places these two vectors differ.

$$\begin{array}{c} \text{FTest} = \{1, 3,\ 1, \boxed{4, 3, 5, 2}\} \ldots\ldots\ldots (1) \\ \text{FTrain} = \{1, 3, 1, \boxed{1, 4, 9, 4}\} \ldots\ldots\ldots (2) \\ \hline \text{Hamming Distance (HD)} = 4 \end{array}$$

$$\text{Match score} = \frac{7-4}{7} \times 100 // \text{Matching score} = \frac{\text{FeatureLength} - \text{HD}}{\text{FeatureLength}} \times 100$$

Which is $3/7 \times 100 = 23.34\%$. The match score shows the possibility of match with the inside database on the basis of given test input. We can put a threshold here to eliminate some of the unwanted results. It will display only those records whose Matching score ≥ threshold (here the threshold is 14.28).The proposed system is modelled on the basis of seven facial and three anthropomorphic features. The anthropomorphic feature set consists of features like gender, age and height while facial feature set includes values such as face type, face tone, face mark, eye type, eyebrows type, nose type and lip shape. The facial features are considered to be vague and hence they have lower significance. The similarity

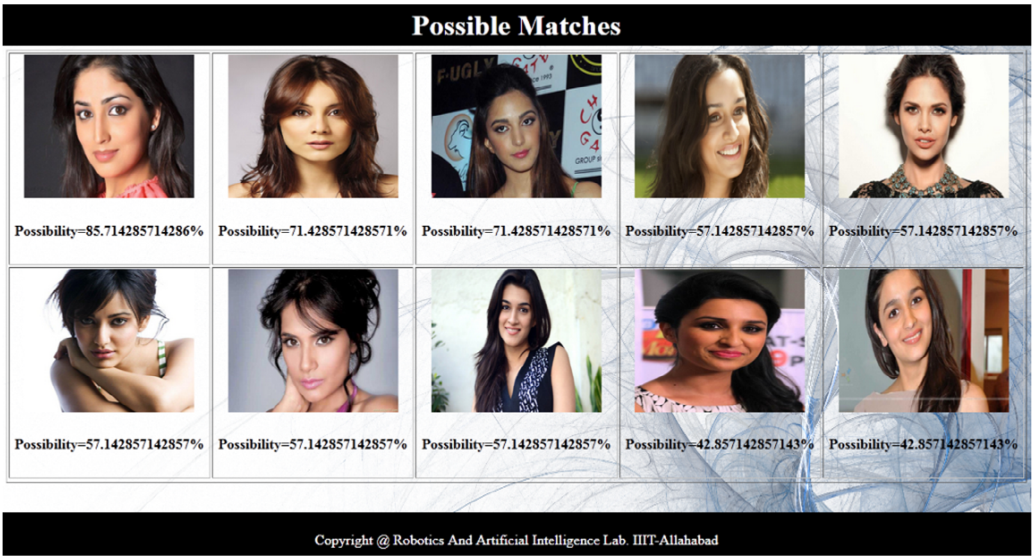

Figure 12. Matched results from the labelled database.

score is computed using only these seven facial attributes. So, if all the attributes are matched, the accuracy would be 100% and if only one is matched the minimum accuracy would be 14.28%. We can define any values between 14.28 and 100 which will act as threshold. In the example shown in this paper, we have used the minimum value as a threshold.

The above figure represents an instance of query execution. Relating to the above figure, let's have a situation where we have to search a Bollywood actress presented in Figure 1. The attributes and their values are already presented in the first example of Section 3. The query optimised by first discriminating males to female. In our database we have total 97 females (57 female celebrities and 40 female students) out of 240 populations. As we can see, the first attribute reduced the search space from 100 to 40%. The next parameter age, again reduces this search span. We have 63% population of the class female which have the average age between 20 and 30. After applying the age filter, we were left with only 15% search space. The height factor further reduces this search span by putting height constraint. At last whatever the population we had, we apply hamming distance between the train and test samples to find the match score. The records were then sorted based on their match score. We have presented the records in the form of top 10, top 20 and top 30. The result of the same query and for the given test face (please refer to Figure 1) is shown in Figure 12.

It could be also possible that during the process of knowledge capturing, eyewitness is not sure about some of the facial as well as anatomical features of the criminal. In order to overcome this, we have introduced the notion of trusted features. Trusted features are those features in which eyewitness is more confident than other features. In such circumstances, the decision tree is generated based on only the trusted features. Suppose the eyewitness is only sure about criminal's gender, age, height, face and cut mark features, then the tree is constructed only using these features.

5. Results and discussion

We have designed two test cases analogous to the viewed and forensic sketches (Klare et al. 2011). In the first test case, the photograph of the guilty is visible to the eyewitness and he/she can describe the attributes on the basis of presented photograph. In the second test case (presented in Figure 13), the photograph of the guilty is absent and was shown one day before the test. Based on these test cases, the matching performance and the validity of the proposed system have been tested. We took help from our robotics and artificial intelligence lab's students to act like eyewitness and to provide the description of these pseudo criminals. We have invited five views per criminal for both the test cases (viewed and forensic). Total 80 students participated in our pen test and each has contributed the views for 15 criminals. We have created a database of 1200 views per test case. The aim of introducing query optimisation is to reduce the search space. We have considered three testing parameters (a) size of search space (b) match score and (c) false negative (FN) to evaluate the performance of our proposed system. It is desired that the search span is minimum while the system shows the maximum match with the respective face in the database. FN are those cases where criminal is present inside the database but the system is not including it inside the search space. This shows the failure of the system. Here we are not concerned about the false positive, as these results are discarded by the eyewitness. The results are decomposed in the form of top 30 matches ranked with respect to their possibility (please refer to Figure 14). The top 30 identities are covered to minimise the chance of FNs. The results on both the test cases are summarised in Figure 14.

There are three abbreviations used in Figure 14, top 10, top 20 and top 30. The meaning of top 10 is that the matched face falls into the first 10 records displayed to eyewitness. Top 20 and top 30 represents the next 10 records after the top 10 and top 20, respectively. A FN of 2.3 and 3.8% has also been reported in test case-I and test case-II.

It is a good sign that in both the cases top 20 and top 30 the performance of test case-II supersedes the performance of test case-I but when we see the overall scenario the aggregated results overcame the FN in test case-I than test case-II. The results are summarised in Tables 11 and 12.

Table 11 is showing the result analysis for the top 20 best matches while Table 12 shows the same analysis on top 30 results. There are three categories of population in our experimental study. The first and second column of each table is representing the different categories and views (population) per category. The third and fourth column show how many views are correctly classified with respect to top 20 and top 30 results. Their individual accuracies are reflected in fifth and sixth column. If we see the individual class accuracy for both the cases, the accuracy of second test case is better than the first case. In first case when photograph (test photograph) is displayed to the user, it's not the same photograph (train photograph) which is used by the expert. So, if the user is not very much familiar to the test face, he/she could misjudge the facial attributes which will cause misclassification. Whereas, if the same person's virtual image is stored in the user long-term memory, he/she can provide better details just after recalling the virtual details. Thereby, the accuracy is increasing in the second case. Another reason which supports the aforementioned analysis is the image used in test case-I is a 2D representation of

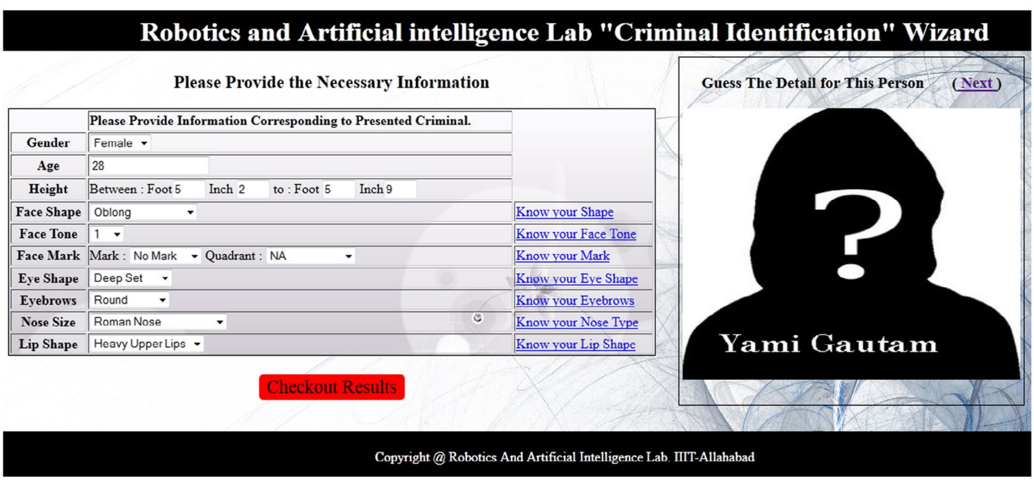

Figure 13. Test case-II (Photograph of the person is absent only caption is available).

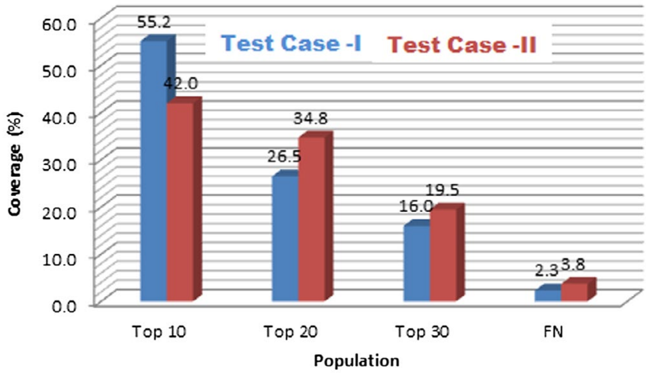

Figure 14. Performance evaluation under both the test cases.

Table 11. Top 20 result analysis.

Categories		Views classified		Accuracy (%)	
	Views per class	Case-I	Case-II	Case-I	Case-II
Bollywood male	265	42	76	15.84906	28.67925
Bollywood female	285	66	87	23.15789	30.52632
Cricketers	200	53	74	26.5	37
IIIT male	250	86	97	34.4	38.8
IIIT female	200	71	83	35.5	41.5
Total	1200	318	417	26.5	34.75

Table 12. Top 30 result analysis.

Categories		Views classified		Accuracy (%)	
	Views per class	Case-I	Case-II	Case-I	Case-II
Bollywood male	265	18	29	6.792453	10.9434
Bollywood female	285	30	43	10.52632	15.08772
Cricketers	200	26	33	13	16.5
IIIT male	250	63	68	25.2	27.2
IIIT female	200	55	61	27.5	30.5
Total	1200	192	234	16	19.5

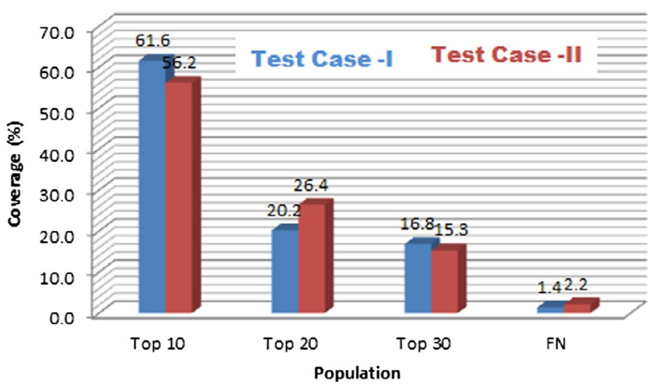

Figure 15. Results after including two additional features.

Table 13. Spread sheet analysis of the result obtained.

Bollywood male: total population 53			
	Age (20–30)	Age (31–40)	Age(41–50)
Samples	12	10	31
Views	60	50	155
	Total views: 265, contribution: 22.08%, classification: 100%		
Bollywood female: total population 57			
	Age (20–30)	Age (31–40)	Age(41–50)
Samples	21	26	10
Views	105	130	50
	Total views: 285, contribution: 23.75%, classification: 100%		
Indian cricketers: total population 40			
	Age (20–30)	Age (31–40)	
Samples	22	18	
Views	110	90	
	Total views: 200, contribution: 16.67%, classification: 100%		
IIITA database: total population 90			
Female	Age (20–30)	Male	Age (20–30)
Samples	40	Samples	50
Views	200	Views	250
Total views: 200, contribution: 16.67%		Total views: 250, contribution: 20.84%	
Test case-I	Classification: 17.20%	Classification: 18%	
Test case-II	Classification: 17.70%	Classification: 16.10%	
Total accuracy: 35.20%		Total accuracy: 33.80%	
After introducing additional attribute			
Test case-I	Classification: 17.25%	Classification: 18.84%	
Test case-II	Classification: 19.25%	Classification: 16%	
Total accuracy: 36.09%		Total accuracy: 35.25%	

the object while the virtual perception about the same object has the same representation in the 3D space which gives the clear and accurate idea about facial and anthropomorphic features. Another difficulty which a user can face in the first test case for both the Bollywood celebrities and Cricketers is about judging their age and height from the photograph, while if the user is recalling the same from the virtual representation. They get more clear idea about the height and age of the suspect.

These results are further refined by two additional factors. The first factor is the measurement of approximation for age and height, the second factor introduces the concept of trusted features. The close approximation about age and height helps in setting up the hierarchy of records while the trusted features gives you the freedom to work on only certain feature set. The approximation is evaluated by measuring the gap between the actual age and height presented by the eye witness and the extracted age and height from the labelled database. We have defined the range of ± 5 year for age and ± 7.62 cm for height attribute. The value of approximation is computed by:

$$\text{Age}_{\text{Diff}} = abs(\text{Age}^{\text{actual}} - \text{Age}^{\text{retrived}} \tag{1}$$

$$\text{Height}_{\text{Diff}} = abs(\text{Height}^{\text{actual}} - \text{Height}^{\text{retrived}}) \tag{2}$$

$$\text{Age}_{\text{Approximation}} = \frac{\text{Age}_{\text{Diff}}}{\text{Age}_{\text{Intervl}}} \tag{3}$$

$$\text{Age}_{\text{Approximation}} = \frac{\text{Age}_{\text{Diff}}}{5}$$

$$\text{Height}_{\text{Approximation}} = \frac{\text{Height}_{\text{Diff}}}{\text{Height}_{\text{Intervl}}} \tag{4}$$

$$\text{Height}_{\text{Approximation}} = \frac{\text{Height}_{\text{Diff}}}{7.62}$$

The fetched records are arranged on the basis of their age and height approximation. Further the matching is performed only on the selected trusted features. Trusted features are those features on which eyewitness is sure. Like if he/she agrees on face type, face tone and eye shape then the matching will be performed only on these features. The effects of using two features are summarised in Figure 15.

The results are very surprising after introducing the additional two features. The FN is minimised and reached up to 1.4 and 2.2% with respect to both the test cases. We have presented a spread sheet analysis of the results obtained from our experimental evaluation in Table 13. The entire population of 1200 views is divided separately over 285 Bollywood female views, 265 male views, 200 Indian cricketer's views, 250 and 200 male and female population views. It has been found that all those views which belong to Bollywood male and female population, Indian cricketer's population are classified with the accuracy of 100%. Their contributions towards the total population views are 23.75, 22.08 and 16.67%, respectively. In total they cover 62.50% of the total population and total accuracy. We have achieved these kinds of fine results because the number of samples for each individual age group is less than the coverage ratio (top 30 results). But in case of our in house database we have achieved only 35.20 and 33.80% of accuracy with respect to test case – I & II. The fall of accuracy is due to the large number of population belonging to the same age group 20–30 and average height of 5 ft 5 inch. The improvement in accuracy is noticed after the addition of two new parameters (a) closure approximation and (b) trusted feature concept. The classification accuracy of male students was improved from 17.16 to 17.25% for test case-I and 18 to 18.33% for test case-II. The same happened with female population the classification accuracy reached up to 19.25% from 17.67% with respect to test case-I, while suffers little bit in test case-II. If none of the feature matched with the existing labelled database, it is very hard to produce accurate results. The problem becomes more severe when the majority of people belong to the same partition. As happened in our in house database case.

5.1. Mug shot vs. proposed approach

If we closely observe the process of mug shot detection, it consists of two major steps. The first step is the pictorial representation of eyewitness perception and the second is to design an algorithm which maps these sketch images to the corresponding photo images of the criminal. To design a sketch similar like the actual criminal is a challenging task which needs proper guidance from the eyewitness. On the other hand, the algorithm designed for matching the sketch images to the existing database should have the calibre and capacity to handle the non-linearity in sketch images. This non-linearity is due to the different kinds of noise (absence/presence of beard, different hair style, shading to show 3D face structure) present in the sketch images. In comparison to the mug shot detection, the proposed approach captures the same perceptual information of criminal's face and physiological structure in the form of symbols and attributes. Further, these symbols are processed with the help of query processing module. The results depicted in Figures 14 and 15 clearly show the strength of this approach. The proposed system covers around 98.6 and 97.8% in both the test cases analogous to view and forensic sketches (Klare et al. 2011). In comparison to the mug shot detection approaches presented in (Tang & Wang, 2003; Uhl & da Vitoria Lobo, 1996; Xiao et al. 2009), the proposed approach performs well in those scenarios where the eyewitness perception is mature. Searching cost of the proposed system is also very low. It reduces the search space to 2.5% (top 30) of the entire search space. The total process takes an average time of 2–4 min which emphases to use this module at the very initial stage of investigation.

5.2. Discussion

We had two test cases in our experiment. Based on these test cases we have reported different results. We figure out the reason behind these differences, one of the possible reasons suggested by Geiselman, Fisher, MacKinnon, and Holland (1985) is the way we are retrieving knowledge from the eyewitnesses about the suspect. The quality of information extracted from the eyewitness depends on the media which is established to communicate with him/her. Geiselman has demonstrated three media of communication established with the eyewitness for gaining the valuable information from him. These three interview procedures are (a) cognitive interview – the knowledge is extracted with the help of some memory retrieval devices (polygraph test), (b) hypnosis interview – to know about the present mental status and (c) standard police interview. Both the cognitive and hypnosis interview show greater efficiency than the standard interview for extracting valuable clues. The other factor which could also affect the identification accuracy is due to the different level of perception maturity of eyewitness. 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). Stronger the perception more accurate the result is. Further, we were interested in how human mind stores the information about any stimulus. How long it is stored in the brain and at which rate it vanishes from the brain. Our brain has three different parts for memory. These are perception, attention and storage of the information. These parts and their connectivity are depicted in Figure 16. Special senses are involved in case of perception, while Thalamus, Frontal lobe & prefrontal cortex are responsible for attention. Hippocampus & Medial temporal lobe stores information. Memory is the ability to encode store & recall an event. Processes involved in human memory are therefore encoding, storage & recall or retrieval (Koriat, Goldsmith, & Pansky, 2000). There is a fourth – memory consolidation which could be 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 & dress & vehicle pass unnoticed. There are also three different kinds of memory we have in our brain listed in Table 14. This clearly shows that if the eyewitness had the criminal face or physical information in his long-term memory, it will produce better results.

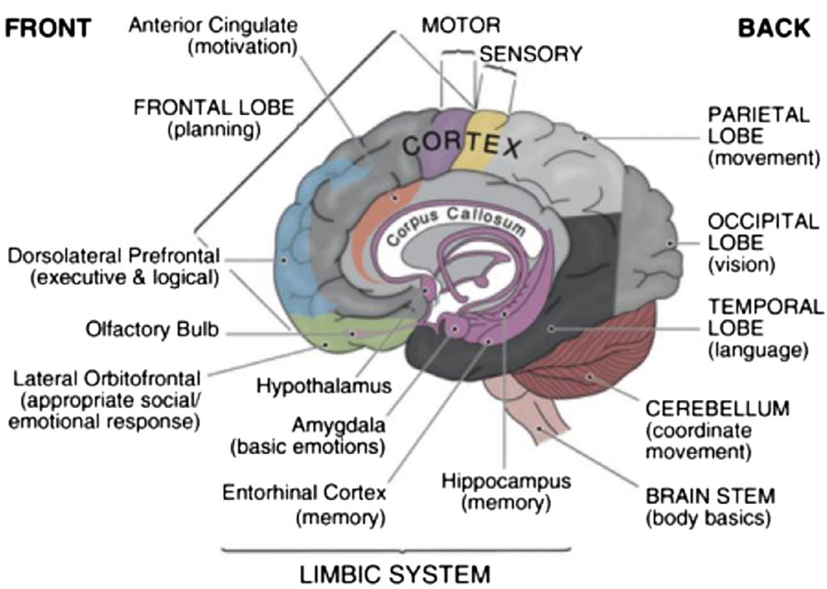

Figure 16. Human brain and internal processing (Koriat et al. 2000).

Table 14. Memory type and store duration.

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

6. Conclusion

A new technique has been introduced in this paper in order to minimise the manual search cost of criminal identification. Two case studies have been considered to test the effectiveness of the proposed approach. In first case, the photograph of the criminal is available to the eyewitness in the time of knowledge acquisition whereas, in the second case, these are absent (shown one day before). The performance of the system has been tested on 1200 views taken from both test cases. In first case study, we have achieved encouraging results which describe that 55.2% of the entire population falls into top 10 matches whereas 26.5% and 16% of the population belongs to top 20 and top 30 matches, respectively. This study infers that the search space has been reduced up to only 2.5%. The results obtained from second case study demonstrates the reliability of this approach with the accuracy level of 42, 34.8 and 19.5% of the entire population on top 10, top 20 and top 30 matches. We have also reported the FN of 2.3 and 3.8% with respect to first and second case study. Further, these results are tuned by introducing the concept of closure approximation and trusted attributes. The closure approximation helps to index the searched records while trusted attributes help to compute the match score only on these attribute set. These two additional attributes minimise the FN to 1.4 and 2.2% in both the test cases and increases the level of accuracy up to 61.6 and 56.2% for top 10 matches.

The future work includes that the additional attributes like weight, hair style, eyeball colour etc. can be incorporated to produce more correct matches. As a future scope, this module could be applied prior to using mug shot detection module. The strength of this proposed work implies that only user eyewitness's notions are required to identify the criminal which avoids the complexity of making the sketch.

Acknowledgements

Author would like to thank Dr Vasudev Joshi former doctor at Banaras Hindu University (BHU), 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 Miss Neha Baranwal, Mr Anup Nandy and Mr Akshay Singhai for assisting over this project.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Chellappa, R., Wilson, C. L., & Sirohey, S. (1995). Human and machine recognition of faces: A survey. *Proceedings of the IEEE, 83*, 705–741.

Eye Shape. (2014). Retrieved from http://www.beautylish.com/a/vcacj/whats-your-eye-shape

Gao, X., Zhong, J., Tao, D., & Li, X. (2008). Local face sketch synthesis learning. *Neurocomputing, 71*, 1921–1930.

Geiselman, R. E., Fisher, R. P., MacKinnon, D. P., & Holland, H. L. (1985). Eyewitness memory enhancement in the police interview: Cognitive retrieval mnemonics versus hypnosis. *Journal of Applied Psychology, 70*, 401–412.

Gibson, L. (2010). *Forensic art essentials: A manual for law enforcement artists*. Burlington, MA: Academic Press.

Gudivada, V. N., Raghavan, V. V., & Seetharaman, G. S. (1993). An approach to interactive retrieval in face image databases based on semantic attributes. In *Third Annual Symposium on Document Analysis and Information Retrieval*. Las Vegas, NV: University of Nevada.

Hamming Distance. (2014). Retrieved from http://www.maths.manchester.ac.uk/~pas/code/notes/part2.pdf

Human Anatomy. (2014). Retrieved from http://design.tutsplus.com/tutorials/human-anatomy-fundamentals-advanced-facial-features--cms-20683

Jain, A. K., Klare, B., & Park, U. (2011). Face recognition: Some challenges in forensics. In *IEEE International Conference on Automatic Face & Gesture Recognition and Workshops* (pp. 726–733). Santa Barbara, CA: Fess Parker's Doubletree Resort.

Jain, A. K., Klare, B., & Park, U. (2012). Face matching and retrieval in forensics applications. *IEEE Multimedia, 19*, 20–28.

Klare, B. F., Li, Z., & Jain, A. K. (2011). Matching forensic sketches to mug shot photos. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 33*, 639–646.

Jarke, M., & Koch, J. (1984). Query optimization in database systems. *ACM Computing Surveys, 16*, 111–152.

Klare, B. F., Klum, S., Klontz, J. C., Taborsky, E., Akgul, T., & Jain, A. K. (2014, September). Suspect identification based on descriptive facial attributes. In 2014 IEEE International Joint Conference on Biometrics (IJCB) (pp. 1–8). Florida, FL: IEEE.

Koriat, A., Goldsmith, M., & Pansky, A. (2000). Toward a psychology of memory accuracy. *Annual Review of Psychology, 51*, 481–537.

Kuei, C. A. (1998). *Face reading*. Lanham, MD: M. Evans and Company. ISBN:087131875X, 9780871318756.

Liu, Q., Tang, X., Jin, H., Lu, H., & Ma, S. (2005). A nonlinear approach for face sketch synthesis and recognition. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Vol. 1, pp. 1005–1010). San Diego, CA. Nice, France.

Luschan, V. (2014). Retrieved from http://en.wikipedia.org/wiki/Von_Luschan%27s_chromatic_scale

Magerman, D. M. (1995). Statistical decision-tree models for parsing. In *Proceedings of the 33rd Annual Meeting on Association for Computational Linguistics* (pp. 276–283). Stroudsburg, PA.

Mane, D. R., Kale, A. D., Bhai, M. B., & Hallikerimath, S. (2010). Anthropometric and anthroposcopic analysis of different shapes of faces in group of Indian population: A pilot study. *Journal of Forensic and Legal Medicine, 17*, 421–425.

Siddiquie, B., Feris, R. S., & Davis, L. S. (2011, June). Image ranking and retrieval based on multi-attribute queries. In *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 801–808). Providence, RI: IEEE.

Tang, X., & Wang, X. (2002). Face photo recognition using sketch. In *International Conference on Image Processing Proceedings* (Vol. 1, pp. I–257). Rochester, NY.

Tang, X., & Wang, X. (2003). Face sketch synthesis and recognition. In *IEEE International Conference on Computer Vision Proceedings* (pp. 687–694). Nice, France.

Uhl, R. G., Jr., & da Vitoria Lobo, N. (1996). A framework for recognizing a facial image from a police sketch. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 586–593). San Francisco, CA.

Wang, X., & Tang, X. (2009). Face photo-sketch synthesis and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 31*, 1955–1967.

Wells, G. L. (1978). Applied eyewitness-testimony research: System variables and estimator variables. *Journal of Personality and Social Psychology, 36*, 1546–1557.

Xiao, B., Gao, X., Tao, D., & Li, X. (2009). A new approach for face recognition by sketches in photos. *Signal Processing, 89*, 1576–1588.

Xiao, B., Gao, X., Tao, D., Yuan, Y., & Li, J. (2010). Photo-sketch synthesis and recognition based on subspace learning. *Neurocomputing, 73*, 840–852.