CRIMINAL IDENTIFICATION SYSTEM BASED ON FACIAL RECOGNITION USING GENERALIZED GAUSSIAN MIXTURE MODEL.

Uttam Mande*, Y.Srinivas², J.V.R.Murthy

1Dept of CSE, GITAM University, Visakhapatnam, India.
2Dept of IT, GITAM University, Visakhapatnam, India.
3Dept of CSE J.N.T.University, Visakhapatnam, Kakinada, India.

ABSTRACT
There is an abnormal increase in the crime rate and also the number of criminals is increasing, which leads towards a great concern about the security issues. Crime preventions and criminal identification are the primary issues before the police personnel, since property and lives protection are the basic concerns of police but to combat crime, the availability of police personnel is limited. To help the cops, comprehensive data regarding the criminals will be advantageous. Data mining concepts proved to yield better results in this direction. In this paper, binary clustering and classification techniques have been used to analyze the criminal data. The crime data considered in this paper is from Andhra Pradesh police department this paper aims to potentially identify a criminal based on the facial evidence obtained through the CCT cameras or the identification based on witness/clue at the crime spot using a Generalized Gaussian Mixture model.

INTRODUCTION
Undoubtedly, there are significant changes in the living styles of humans, which may be due to the effect of technological growth and modernization or environmental conditions that force the humans to indulge in criminal activities. Law keepers are working rigorously to maintain law and order, but at the same time, the crime activities are increasing enormously leaving the cops difficult to analyze the crime and arrest the criminals [1][2]. Lot of research is projected in this direction by construction different models by the researchers for effective analysis [3][4][5][6]. However, the main disadvantage is that due to the overload of data regarding the crime activities, together with the increase in the number of criminals makes it difficult in analyzing the data. Therefore a better model with the knowledge about the crime & the criminal always will always be advantageous. Data mining techniques will be very much useful for these purposes [5][7][8] for analyzing enormous data.

The usage of data mining helps in exploring the voluminous data and help the law keepers to retrieve the information more effectively and efficiently. With the usage of the data mining concepts, such as clustering and classification analyzing the large data gets simplified and this helps in the identification of the criminals. In order to identify a criminal, either direct evidence from the witness or indirect witness gathered by the forensic experts from the spot will be of great use. In this paper, we use the information available at the crime spot.

The faces which are recorded in the CCT cameras are considered as feature vectors and using the features from the face, the identification of the criminal is carried out. Generally the facial recognition systems operate in two modes, open system and closed system. In closed system, the test face to be compared should be one among the faces in the database, and in the open system, also called imposter, the faces in the database and the test face will be different. Most of the facial recognition algorithms belong to either discriminative group or generative group Kieron Messer et al., (2004) [9]. The identification of the faces in the discriminative approaches, also called non-model based, (PCA, LDA etc) is done based on true/ false matching, and hence they are also termed as holistic approaches. Generative approaches also known as model based approaches, are considered to be more efficient than discriminative approaches, and most of the latest face recognition systems are developed using Gaussian mixture models [5,18,22]. The model based approaches for automatic face recognition systems, are more effective than the non-model based, since in the earlier method, we consider the parameters of the faces into considerations for the identification. In model based face recognition systems, the feature vector is characterized through a probability distribution.

The primary step in the identification of any facial recognition system is feature vector extraction. In order to extract the feature vectors, many researchers have used...
either 2D Discrete Cosine transformation or DCTmod2 or 2D Gabor Wavelets [10, 11, and 12]. In these methods, they have considered that the feature vectors extracted by DCT follow a Gaussian mixture model. But in general, most of the faces in reality are either meso-kurtic, platy kurtic or symmetric, and the Gaussian mixture model based facial recognition systems perform well only if the facial image under consideration is symmetric. Hence to consider the faces, both symmetric and non-symmetric faces, we need to use more general models. Hence in this paper, the Generalized Gaussian mixture model is used for modeling the feature vectors. The rest of the paper is organized as follows, section -2 of the paper deals with deep insight into fundamentals of crime analysis, in section -3 the concept of binary clustering is presented, the Generalized Gaussian Mixture model is discussed in section -4, experimentation is highlighted in section -5, the section 6 of the paper focus on the conclusion.

1. Features for criminal identification

Any crime investigation is based on the either the direct witness or indirect evidence. Direct evidence / witness help in the proper identification of the criminal. The clues obtained in the form of witness help the stepping stone towards the crime analysis, and criminal relating is the mapping of the criminal based on the clues with data available in the data base, by the use of intelligent knowledge mapping.

The crime data base considered in this papers includes the criminals involved/accused in several types of crimes such as 1) robbery 2) murder 3) kidnapping 4) riots

1 crime Variables

The primary crime variables that would be considered for crime investigation are:

1) Crime Spot (place: restaurant, theater, road, railway station, shop/gold shop, mall, house, apartment )
2) Criminal attribute(hair, built, eyebrows ,nose ,teeth, beard , age group,mustache, languages known)
3) Criminal psychological behavior can be recognized by type of killing

We have considered the type of killing as (smooth, removal of parts, harsh) which attributes to the psychological behavior of the criminal

4) Modus operandi (object used for crime), 1)Pistol 2)Rope 3)Stick 4)Knife

These criminal links help to analyze the dataset there by making the crime investigators to plane for identification of the criminal.

2. Clustering:

In order to simplify the analysis process the huge dataset available is to be clustered. The clustering in this paper is based on the type of crime. A data set is generated from the database available from the Andhra Pradesh police department and a table is created by considering the FIR report.

The various fields considered including the criminal identification numbers, criminal attributes, criminal psychological behavior, crime location, time of crime (day/night), witness /clue, the data set is generated by using the binary data of 1’s & 0’s, 1’s indicating the presence of attribute and 0’s indicating the absence of attribute then clustering of the binary data is done as proposed by Tao Hi (2005) using the binary clustering.

Crimes are categorized in many ways, here we have given weights to each type of crime where weighing scheme is considered in the manner all the relative crimes will be given with near values, after applying clustering algorithm on this type of crime feature we have got four clusters of crime data they are robbery, kidnap, murder and riot.

Fig 1 categories of crimes

3. Generalized Gaussian Mixture model

The Probability Density function the Generalized Gaussian mixture model is given by

\[
f(z | \mu_k, \sigma, P) = \frac{1}{2\Gamma(1+\frac{1}{P})A(P, \sigma)} e^{-\frac{\left[Z(z_0) - \mu_k\right]^P}{A(P, \sigma)}}
\]

\[
\sigma > 0, A(P, \sigma) = \left[\frac{\sigma^2}{\Gamma(\frac{1}{P})}\right]^{\frac{1}{2}} \Gamma\left(\frac{3}{P}\right)
\]

The initial estimates are updated using the EM algorithm and the final updated parameters are given by

\[
\mu_k^{(l+1)} = \frac{\sum_{x=1}^{N} t_i(z_s, \Theta^{(l)}) \gamma^{(N,P)}z_s}{\sum_{x=1}^{N} t_i(z_s, \Theta^{(l)}) \gamma^{(N,P)}}
\]

\[
\mu_k^{(l+1)} = \frac{\sum_{x=1}^{N} t_i(z_s, \Theta^{(l)}) \gamma^{(N,P)}z_s}{\sum_{x=1}^{N} t_i(z_s, \Theta^{(l)}) \gamma^{(N,P)}}
\]

In order to identify the criminals the variables/links that are identified from section 2.1 are mapped to that of the data base and previous knowledge there by solving the crime incidents, in order to identify a criminal together with crime variables that are discussed in section 2.1 we have considered 1) witness available and 2) clues available

The various witness considered for affective identification are discussed in the above section if the evidence/witness is not available we have considered clues available from the forensics such as finger prints in particular cases for the identification we have considered the mapping of both witness & clues to authenticate a criminal.

In this paper we have considered clustering to cluster the data base based on type of crime and the classification is carried out from the feature available.
In this paper we have considered the feature called the witness available from the security cam or multiple facial images available from digital cams/camera’s available/the witness who witnessed the crime. The identification of a face is also carried out from multiple objects retrieved from sources. Extracting faces, Comparing data bases and the outputs generated are presented in the figures 3, 4, 5 and 6. Each of the images retrieved from the CCT camera, we generate a PDF using the Generalized Gaussian Mixture model and for all the images in the Database, the PDFs are calculated and stored. If a match exists between the test PDF and the PDF in the database, the relevant facial image with the PDF is retrieved and the information regarding the criminal is obtained. Another variation in this paper, is that extracting a facial image by co-relating with the features described the witness and generating the image that suits the features. The features obtained from the input sources are compared and these features are given as the inputs to Generalized Gaussian Mixture Model The PDFs of each of these face using feature are calculated and nearest faces are identified

$$\sigma_i^{1+p} = \left[ \frac{N}{\sum_{i=1}^{N} l_i (z_i, \theta^{(i)}) \left( \frac{\Gamma(3/P)}{P T(1/P)} \right) z_i - \mu_i^{(i)} \right]^{1/2}$$

$$\mu_i^{(i+1)} = \frac{\sum_{x=1}^{N} l_i (z_x, \theta^{(i)}) z_x}{\sum_{x=1}^{N} l_i (z_x, \theta^{(i)})}$$

4. Experimentation
The experimentation is done in Mat lab environment. The data base is generated with the following data

<table>
<thead>
<tr>
<th>Criminal Id</th>
<th>Type of crime</th>
<th>Mode of Crime</th>
<th>Place</th>
<th>Shop</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>102</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>103</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>104</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>106</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>107</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>108</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>109</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>110</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>111</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>112</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>113</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>114</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>115</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>116</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>117</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>118</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>119</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>120</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>121</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>122</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>123</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
These features are used to generate a image. The images so generated is stored and can be used. The features obtained from the input sources are compared and these features are given as the inputs to New Gaussian Mixture Model. The PDFs of each of these face using feature are calculated and nearest faces are identified.

CONCLUSION
If the witness is available, at the crime incident, then this paper brings out a novel methodology of constructing a image with the features and comparing with the faces in the database and if a match is obtained, it tries to present the details regarding the criminal and his identification marks. This paper also helps in situations where the face of a criminal is obtained from a CCT camera and this face will be compared with that of the faces in the database to ratify a criminal. In these situations, in this paper we have tried to identify the criminal by mapping the criminal using the Generalized Gaussian mixture model.

REFERENCES
2. Cate H. Fred “Legal Standards for Data Mining” retrieved from the internet on 12-03-2011
http://www.hunton.com/files/tbl_s47Details/FileUpload265/1250/Cate_Fourth_Amendment.pdf
4. Jeff and Harper, Jim “Effective Counterterrorism and the Limited Role of Predictive Data Mining” retrieved from the web 12-02-2011