Hidden Markov Model for Identification of Different Marks on Human Body in Forensic Perspective

Dayanand G Savakar
Department of Computer Science, Rani Channamma University, Belagavi India
Email: dgsavakar@gmail.com

Anil Kannur
Department of Computer Science & Engineering, VTU Resource Research Center, Belagavi India
Email: anilkannur1978@gmail.com

Received: 15 January 2019; Accepted: 13 February 2019; Published: 08 March 2019

Abstract—This paper proposes a computational forensic methodology which identify and classify different marks on the human body using Hidden Markov model. The methodology gives an efficient and effective computerized approach for the characteristics of different marks such as birthmarks, burntmarks, tattoos and weapons’ wounds found on human body. This proposed method will be a computationally effective substitution for the traditional forensic method in identifying the body marks in crime investigation of homicidal cases. Hidden Markov Model (HMM) is statistical and logical tool suitable for this identification. The marks on human body describe different patterns with characteristics that are helpful in identification. The experimental results achieved for identification of different marks with an average accuracy of 94.6%, on the available database of 400 images that includes four categories: Birthmarks, Burntmarks, Tattoos and weapons’ wounds (100 images of each marks). The methodology gives the better combination of features (color, texture and shape), which are extracted for the identification of marks on human body for the purpose of computational forensic science.

Index Terms—Birthmark, Burntmarks, Hidden Markov, Identification, Segmentation, Tattoos, Weapons, Wounds

I. INTRODUCTION

Hidden Markov Models are type of statistical models accustomed characterize the noticeable properties of a representation. HMMs incorporates two interrelated processes, firstly an underlying, unobservable Markov chain with finite number of states rule by a state likelihood matrix and an initial state likelihood distribution and secondly a set of observations, outlined by the observations density functions related with each state. In this paper, a new method to identify the different marks on the human body is proposed based on the HMMs. HMM is an extensive method to likelihood sequence modeling, viewed as stochastic simplifications of finite state machines, where transitions between states and generation of output symbols are ruled by likelihood distributions (Rabiner, 1989; Baum et al., 1970). HMM is used for various image processing and pattern identification applications; among these applications, not many of the researchers worked on the different marks on the human body, these marks definitely have characterized patterns which will be helpful in identifying, even today traditional method: pathologist or medical practitioner recognize these marks manually but this traditional method has many disadvantages as it includes human error and manipulation of data. To eliminate these disadvantages, the system should be without human. In this paper, a computerized method is proposed to identify different marks on human body, a process model for digital investigations that meets the following requirements: Identification of the different marks into four main categories: Birthmarks, Burnt Marks, Tattoos and Wounds caused by weapons. To achieve better performance results, a statistical tool Hidden Markov Models used. HMM is widely used in various applications and applying HMM in real world application such as this proposed methodology and various other engineering applications will improve the detection accuracy for the whole field [3]. This proposed method will be helpful in forensic investigation. The subsequent section of this article includes related work, problem statement and experimental setup and results with analysis. The adaption of HMM for the identification and classification is that the model works on recovering of data in sequence which will be more suitable for this proposed methodology.

II. RELATED WORK

Many of the researchers have used conventional, medical strategies for identification while others have used heuristic techniques for segmentation followed by...
ANN [17] and fuzzy primarily based techniques for the wound class manner [10, 23]. Another proposed system that robotically and as it should be identify the location of a persistent wound and is capable of computerized picture segmentation and wound location identification [18, 20]. Two distinctive varieties of ANNs, the Multi-Layer Perceptron and the Radial Basis Function with parameters decided via a pass-validation method, are then implemented with supervised mastering inside the prediction system for the wound identity, and their consequences are compared [8]. Other approach to determine particular patterns and distribution of defense accidents performed on 121 homicidal deaths which confirmed protection accidents in 40 cases [21, 22]. The Image evaluation method is used in identifying the styles at the wounds [6, 10, and 12]. The strategies used most commonplace in identification is post-mortem research, an uncommon harm turned into observed within the area of the fatal wound to the neck, suggesting that the wound turned into inflicted with an exceptional pressure, the usage of a pointy item with a weird shape. A comparative examine among the 2 weapons become done to be able to decide the compatibility with the above-mentioned damage. Consequently, those analyses led us to the identification of the murder weapon and the author of the crime [10, 14]. Other researchers proposed widely-wide-spread version and gadget to effectively determine, compare and perceive the foremost guns from a massive series of available weapon structures that have more than one criteria based on a fuzzy choice-making model [17]. These techniques do not employ digital investigation algorithms. Through employing more powerful identity strategies so that you can reduce the human errors, this in

III. PROBLEM DEFINITION

The Literature survey has shown that not much work has done yet in computerized identification of different marks on human body, so this has given a scope of developing methodology to analyze and detect different marks on the human body. Recent development in research has proved that computer vision can be developed feasibly into a right tool to recognize evidences during crime investigations. This paper proposed the methodology which identifies different marks such as birthmarks, burnmarks, tattoos and weapons' wounds which are available on human body. The majority of these works have used characterized images acquired under controlled conditions. The execution of machine vision in detection of body marks in digital images that take care of oblige of calculations that separate examples. This methodology uses Hidden Markov Model for the identification process and the methodology contains two phases: training and testing phase, refer Figure 1.

![Block Diagram of Proposed Methodology](Image)

For the both the phase, acquiring images, segmentation and features extractions remains the same procedure. In training phase, HMM will train with sample images of training datasets stored in Knowledge base. In testing phase, trained HMM is used to identify an unknown sample images which will be acquired in real time. This methodology has considered suitable and better combination of features which are extracted for the analysis and identification process.

IV. CHARACTERISTICS OF THE MARKS

Body marks can be characteristic of an individual and can be used to support an identification, together with scientific or police statistics and with identity given by using own family individuals. Tattoos are patterns at the skin fashioned via injection of dyes right into a pattern of prick marks. The presence of tattoos can characterize a number of meanings. Birthmarks are some other essential frame mark. Birthmarks are benign tumors regarding the blood vessels simply underneath the skin. They are often present at birth, because the call suggests, although they will fade with time. Moles and warts are other essential skin capabilities, although they're extra commonplace and less character than a birthmark. Weapons’ wound pattern are another important type of frame mark. Wound type can also vary from bullet, stabbing, blunt pressure, rape, poisoning, burn, or site visitors’ fatality.
V. IMAGE ACQUISITION AND PREPROCESSING

The images were acquired with color Digital camera under controlled conditions. The camera with a zoom lens 16x focal length and close-up set 1x. The Images were acquired with higher resolutions in size (refer Figure 2 for examples). The controlled condition indicates that there will not be any further damages made to victims’ body and at specially marks or wound region. The quality of image is pivotal for further analysis and identification.

That may affect the ability to identify features and objects in image. To enhance the quality of the images, the proposed methodology uses spatial filtering and image sharpening so as to improve the image free from noise and for clear edge detection respectively which will be helpful for segmentation process.

A. Spatial Filtering

Filtering is an essential task in image processing and pattern recognition, used for enhancing the image, reduction of noise, detection of edge and sharpening. Spatial filtering is performed using convolution function as given in Eq. 1

\[ s(x, y) = \sum_{m=-M/2}^{M/2} \sum_{n=-N/2}^{N/2} h(m, n)f(s-m, y-n) \]

where

- \( s(x, y) \) is the filtered image
- \( f(x, y) \) is the original image
- \( h(m, n) \) is the Gaussian filtering mask with size \( M \times N \).

B. Image Sharpening

Image sharpening is an enhancement process that determines the edges and fine details in image[2]. Firstly, the original image is filtered by high-pass filter that acquires the high-frequency components and secondly, the original image is added with scaled form of high-pass filter output, this produces an image with enhancement i.e., sharpening of the original. The sharpening operation can be represented by Eq.2:

\[ S_{i, j} = X_{i, j} + \lambda F(X_{i, j}) \]

where

- \( X_{i, j} \) is original pixel value at \((i, j)\);
- \( F(.)\) is the high-pass filter;
- \( \lambda \) (\( \geq 0 \)) is tuning parameter and the value of \( \lambda \) depends on the degree of sharpness anticipated.

VI. SEGMENTATION

In this proposed methodology, image segmentation is implemented based on the threshold, which is greatly suitable for the image and pattern classification applications [1]. Threshold-based is frequently used algorithm for image segmentation. The function \( g \) of grey value with the threshold is defined by Eq.3:

\[ g(\nu) = \begin{cases} 0, & \nu < t \\ 1, & \nu \geq t \end{cases} \]  

where \( \nu \) is grey value, and \( t \) is threshold value. The function of thresholding is to map grey value image to binary image, where the threshold value is decided interactively based on the users’ manipulation of the value (and also there exist many methods for selecting suitable threshold value) and checking the result, this process is continued till satisfying result is obtained. After the thresholding function, the image is segmented into two; identified by pixel values 0 and 1 respectively.

\[
\text{Algorithm 1: Threshold-Based Segmentation} \\
\text{Input: Captured Color Image} \\
\text{Output: Segmented color Image consists of Region of interest} \\
\text{Start} \\
\text{Step 1: Estimate Initial threshold } T \\
\text{Step 2: } T \text{-based Segmentation: } G_1 \text{ pixels brighter than } T; \ G_2 \text{ pixels darker than or equal to } T. \\
\text{Step 3: Calculating the average intensities } m_1 \text{ and } m_2 \text{ of } G_1 \text{ and } G_2. \\
\text{Step 4: Computing new threshold: } T_{\text{new}} = (m_1 + m_2) / 2 \\
\text{Step 5: If } |T - T_{\text{new}}| > \Delta T \text{ then back to step 2 else stop} \\
\text{Stop} \\
\]

VII. FEATURE EXTRACTION

The region of interest in segmented image is used for extracting the features (refer table 1) for the identification process of different marks such as birthmarks,
burntmarks, tattoos and weapons’ wounds. In this paper, the features of images of said marks were used for identification and to model the listed features, the Markov chain concept is used as a random process. A random process \( \{ O_n, n = 0, 1, 2, \ldots \} \) is called a Markov chain if the probable value in state \( O_n \) depends on the probable value in state \( O_{n-1} \), is given Eq.4:

\[
P(O_n | O_{n-1}, O_{n-2}, \ldots, O_0) = P(O_n | O_{n-1})
\]

The joint probability of a certain observation sequence \( \{ O_1, O_2, \ldots, O_n \} \), using Markov assumption is given by Eq.5

\[
P(O_1, O_2, \ldots, O_n) = \prod_{i=1}^{n} P(O_i | O_{i-1})
\]

We use this Eq.5 to differentiate between the features (Observations) in different images of marks.

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Name of the Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Red Mean</td>
</tr>
<tr>
<td>2</td>
<td>Red Variance</td>
</tr>
<tr>
<td>3</td>
<td>Red Range</td>
</tr>
<tr>
<td>4</td>
<td>Green Mean</td>
</tr>
<tr>
<td>5</td>
<td>Green Variance</td>
</tr>
<tr>
<td>6</td>
<td>Green Range</td>
</tr>
<tr>
<td>7</td>
<td>Blue Mean</td>
</tr>
<tr>
<td>8</td>
<td>Blue Variance</td>
</tr>
<tr>
<td>9</td>
<td>Blue Range</td>
</tr>
<tr>
<td>10</td>
<td>Co-occurrence Matrix</td>
</tr>
<tr>
<td>11</td>
<td>Energy</td>
</tr>
<tr>
<td>12</td>
<td>Contrast</td>
</tr>
<tr>
<td>13</td>
<td>Inverse difference moment</td>
</tr>
<tr>
<td>14</td>
<td>Correlation</td>
</tr>
<tr>
<td>15</td>
<td>Eccentricity</td>
</tr>
<tr>
<td>16</td>
<td>Geometric Moments</td>
</tr>
</tbody>
</table>

The features listed in table 1 are extracted using the algorithm as shown below:

**Algorithm 2: Feature Extraction**
- **Input:** Segmented color Image from the Algorithm 1
- **Output:** Feature Values

**Start**

**Step 1:** Read the Segmented color image
**Step 2:** Separating of RGB components from color image and computation of mean, variance, and range for the components using following equations:

\[
\text{Mean} \mu = \sum_x \sum_y p(x, y)
\]

\[
\text{Variance} = \sum_{x,y} (x - \mu)^2 p(x,y)
\]

\[
\text{Range} = \max(p(x, y)) - \min(p(x, y))
\]

**Step 3:** Computing the following texture features from the image:

**Step 3.1:** Co-occurrence matrix is defined over an image is for a given offset the co-occuring pixel values (gray-scale or colors) distribution. For an image with different pixel values, the \( p \times p \) co-occurrence matrix \( C \) is defined over an \( n \times m \) image \( I \), parameterized by an offset \((\Delta_x, \Delta_y)\), as given in Eq.9:

\[
C_{\Delta_x, \Delta_y}(i,j) = \sum_{x=i}^{n} \sum_{y=j}^{m} \{ \begin{array}{ll}
1 & \text{if } x + \Delta_x, y + \Delta_y = j \\
0 & \text{otherwise}
\end{array} \}
\]

where, 
\( i \) and \( j \) : the pixel values; 
\( x \) and \( y \) : the spatial positions in the image \( I \); and 
\( I(x, y) \) : the pixel value at pixel \((x, y)\).

**Step 3.2:** Energy can be defined as the degree of repetitions of pixel pair at the extent and it gives the image uniformity, is given by Eq.10:

\[
\text{Energy} = \sum_{x,y} I^2(x, y)
\]

**Step 3.3:** Contrast is a pixels’ degree of intensity and its neighbor and is calculated by the difference in objects’ color & brightness and other objects within same field view

\[
\text{Contrast} = \sum_{x,y} |x - y|^2 P(x, y)
\]

**Step 3.4:** Inverse difference moment is defined as degree of local homogeneity of an image and acquires the degree of closeness in distribution

\[
\text{Inverse difference moment} = \sum_{x,y} P(x, y) / |x - y|^2
\]

**Step 3.5:** Correlation is the degree of similarity between data the relates spatial domain processing with frequency domain processing

\[
\text{Correlation} = \sum_{x,y} \sigma_x \sigma_y
\]

\[
\text{Correlation} = \sum_{x,y} (xy)P(x, y) - \mu_x \mu_y
\]

\( \mu_x \) and \( \mu_y \) : means; and \( \sigma_x \) and \( \sigma_y \) : standard deviations.

**Step 4:** Computation of shape features from the image:

**Step 4.1:** Eccentricity is computed using the length of semi-major axis \( a \) and semi-minor axis \( b \) of the object in image and is given by Eq. 14:

\[
E = \sqrt{1 - b^2/a^2}
\]

**Step 4.2:** Geometric moments: If an image is considered as discrete function \( f(x, y) \) with \( x=0,1,\ldots,M \) and \( y=0,1,\ldots,N \), then the geometric moment is given by Eq.15:

\[
m_{pq} = \sum_{x=0}^{M} \sum_{y=0}^{N} x^p y^q f(x,y)
\]

\( p(x, y) \) and \( q(x, y) \) : a non-zero and piece-wise continuous function in finite part of the xy plane (region-of-interest) of the image.

**Stop**
VIII. HMM FOR IDENTIFICATION AND CLASSIFICATION

The process used to acquire the sequence of data from an image contains two steps: reading of image and extracting of features, respectively. In initial step, sequence is acquired by reading the segmented input image. In initial approach, these sequences are obtained using raster scan and is recognized using selective target based scanning. The advantage of using saliency-based approach is two, firstly, relevant information of the image can be target, thus achieving robustness; secondly, as the patches are taken out in decreasing order of importance, this analysis can be stopped based on the decision. Recognition of different body-marks using HMMs usually has solution to the different problems. This recognition process goes in this way: A features sequence is extracted from the segmented input image. To discriminate the systems’ ability with the different body-marks that strongly depends on the kind of features extracted and its performance depends on handling the experimental conditions. The System’s learning problem is resolved by training statistical classifier based on the strategy chosen for classification, in this experimental setup – the classification strategy chosen is mixed classification. The mixed classification are the models that tradeoff between the two methods: training one HMM model for each sequence and one for each class, thus making the classification more complex but efficient. Such mixed approach is dissimilarity-based representation. After the extraction of sequence, the step 2 consists of computation of features from the image patches. Extracted features are examined for the image patches evaluation purpose, for this local neural network models were used, thus generating class-posterior probability vector for each object in the sequence that is used as observation feature in succeeding Markov model. This methodology gives out a better level of classification module range but it is costly to maintain local models in large quantity. The probability computation: A Hidden Markov Model $^\land$ is defined as a set of N emitting states as well as an initial and end of line state, so we end up with N+2 states. The expression $S_t = I$ will indicate the occurrence of state I at time t. The time indices run from $t=1$ to $t=T$, where $T$ is the length of the observation sequence $X = [x_1, x_2, ..., x_T]$ to be matched to HMM. The states are coupled by transition $a_{ij}$ denotes state transition probability with the subscripts indicating the two states involved and $a_{ij}$ refers to the self-loop probability. The first null state has a transition probability of 1 and no self loop probability. Each emitting state has an associated probability density function described as $f_i(x)$ for $s_t \in \land$. A single left to right HMM can now be described as $\land = [a, f]$. The possible solution to this problem is to enumerate all possible sequences of states $S_{T+1} 0$ and determine the value of $f(X T , S_{T+1} 0 | \land )$.

A. Training the Model

The identification and classification process on different body-marks database is summarized as follows: Entire database is first divided into two parts – training and testing database, this training database is used to train HMM models as discussed above, that is each body-mark in the training database, an HMM is trained. Each body-mark test is then counted against all the models, is identified and classified to the model with highest similarity measure. Each count represents the similarity between the test body-mark image and the class and subsequent models to be trained. Every different object in the database is represented by a HMM body-mark model. A set of images is considered that represents different instances, different victim bodies and rotation of same images are used to train each HMM model. In this, features extracted from each segmented image are used to form the observation vectors and these are efficiently used in training of each HMM model. First, the initialization of HMM $\lambda = \{A, B, \Pi\}$ is carried out. The training data is uniformly separated among the N states and the observation vectors related with each state are used to acquire initial estimation of the observed probability matrix $B$. The initial values of $A$ and $\Pi$ are the set given to the structure of body-mark models. These iterations are stopped when the model convergence is achieved.

B. Testing Model of System

After the completion of the learning/training process, each class is mapped to a HMM. For K-class classification problem, there will be K distinct HMM models. Each image in testing database has to pass through feature extraction and quantization process. It is understood that each test image is represented by its own observation vector. In this testing process, for every input image of body-mark from the testing database, there will be calculation of probability of the observation vector (of current test image) given to each HMM model. For a body-mark image I is recognized as body-mark type $X$ if:

$$P(O^{(I)} | \lambda_\chi) = \max_{\lambda_\chi} NP(O^{(I)} | \lambda_\chi)$$  \hspace{1cm} (15)

In order to decrease complexity of computation and consumption of the memory, the proposed method is tested on entire database. The recognition result of the proposed method is calculated for different values of the number of symbols and is changed with the changes happens in levels of quantization. And the list of features extracted and used for the method, has given the best possible and acceptable classification results.

IX. RESULT AND DISCUSSION

The proposed method uses two different databases for training and testing. The database has several instances for each category taken from different human bodies, drawn from a pattern vocabulary of 4 classes. The database is focused on the identification of a vocabulary of different marks of each category. The proposed method has been thoroughly tested on the database which has 100 images for each body-mark making it total of 400 images. This work has focused on the important area of computational forensic science and has obtained
acceptable results that can be compared to research work of same field. The experimental results of applying the proposed methods on various marks on human body has given average accuracy of 94.6%, refer table 2 for results for different marks. As number of hidden states in HMM varies the performance by the model will vary. Identification based on different marks on body parts eliminates the probability of a patterns being predicted as uncorrelated patterns, and thus helps in improving the overall performance of the system. This proposed methodology has given tremendous identification results with set of features chosen. Interestingly, our method is competitive with respect to identification by human experts. In this, where the number of images to identify is so less, and where the perfect identification is required, the method would thus be quite useful. Our experiments show that the method obtains quite good results on available images database representing various wound patterns. Indeed, it is directly applicable to any image identification problem, it is reasonably effective and efficient.

A. Comparison of Result with traditional method

The traditional method for identification is manual investigation by medical/forensic experts, in which comparison study is carried out among images or patterns of different body marks to determine the result. As per the survey, medical/forensic experts recognize different marks with accuracy of 75% to 95%, variations in accuracy of result from one medical/forensic expert to other, as methodology is manual with human intervention so in some of the cases data is manipulated and has risk of human errors wherein computerizing the entire process reduces human errors, manipulation of data can be avoided and can give high accuracy of identification as in this proposed methodology with average accuracy of 94.6%.

Table 2. Identification accuracies for the different marks on the human body

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Birth marks</td>
<td>94.2</td>
<td>82.4</td>
<td>81.7</td>
<td>75% to 92% (Results vary from one Medical practitioner to another)</td>
</tr>
<tr>
<td>Burnt-marks</td>
<td>92.5</td>
<td>75.6</td>
<td>80.5</td>
<td></td>
</tr>
<tr>
<td>Tattoos</td>
<td>96</td>
<td>91.2</td>
<td>87.5</td>
<td></td>
</tr>
<tr>
<td>Weapons’ Wounds</td>
<td>95.7</td>
<td>92.0</td>
<td>85.3</td>
<td></td>
</tr>
</tbody>
</table>

B. Comparison with other Methods

The other methodologies used in identification are image matching technique and artificial neural network, where a comparative study of different body marks was performed in order to determine the result. The results presented by other researchers’ models that recognize different marks has given the average accuracy of 85.3% for image matching technique [19] and 83.8% for ANN model [12], the accuracies were competitively less that has made to propose a new method and proposed method has given high accuracy as obtained with average accuracy of 94.6%.

On the basis of these results, we can say that this approach has considerably improved the identification and classification rate in relation to the image comparison and ANN approach. Moreover, our algorithms allow to detect each body marks and to distinguish it from the others with very good rates with minimum rate at about 92.5% and maximum to 96%. With the HMM algorithm, we obtained an average recognition rate of 94.6%. However, in this approach, the HMM has a better recognition rate. We can say that on this data the HMM is more efficient than the ANN because the recognition rate is improved by more than 10% which is especially not negligible within the framework of body marks detection.

![Comparison of Accuracy](image)

Fig.4. Accuracy Comparison Graphs

X. CONCLUSION

In this article, a novel method was proposed for identifying different body marks using Hidden Markov Models. Threshold based image segmentation is implemented to get the desired region-of-interest. In this method the features selected for extracting are color, texture and shape features of the different body marks, this will be best performing features set to get higher accuracy of identification. HMM is adapted for the identification and classification of the body marks as
involves the recovering data in sequence and has performed better in this proposed methodology. The accuracy for the proposed method has reached 94.25% for the entire testing database, this obtained results indicated that the proposed method had a better accuracy in comparison with traditional method, image matching technique and ANN model. The results of the proposed method show that it is faster and efficient. These results obtained from the proposed method can be used for identifying the different marks on human body in crime and forensic investigations. The proposed method has significance to real world applications and includes image processing and pattern recognition models. And from the literature survey it also shows that very few researchers worked on the said problem statement i.e., identification of different marks as separate study or research when discussing with complete system as proposed in this paper.

REFERENCES


Authors’ Profiles

Dr. Dayanand G. Savakar, Associate Professor, Department of Computer Science, Rani Channamma University, Belagavi, India. He has obtained his B.E in Computer Science & Engineering in 1990, Master’s degree in Software Systems in 1997 and Ph.D. in Computer Science and Engineering in the year 2010. He has published 35 research papers in peer reviewed International Journals and conferences. His research area of interest is Image Processing and Pattern Recognition.

Mr. Anil Kannur, Research Scholar VTU Resource Research Center Belagavi, India. He has obtained his B.E in Computer Science & Engineering in 2001, M.Tech.(Computer Science & Engineering) in 2006 and presently perusing Ph.D. in Computer Science and Engineering (Registered under Visvesvaraya Technological University, Belagavi, India). He has published 10 research papers in peer reviewed International Journals and conferences. His research area of interest is Image Processing and Pattern Recognition.