# The Study of Automatic Authentication of Printed Security Documents

*Biswajit Halder*

# The Study of Automatic Authentication of Printed Security Documents

*Thesis submitted in the partial fulfillment of the requirement for the degree of*

*Doctor of Philosophy*

*in*

*Computer Science*

*by*

Biswajit Halder

*Under the guidance of*

Dr. Utpal Garain
*CVPR, ISI, Kolkata, WB*

&

Dr. Abhoy Chand Mondal
*Univ. of Burdwan, Burdwan, WB*

Department of Computer Science
**The University of Burdwan**
**Burdwan - 713 104, India**

**July 2015**

# CERTIFICATE

––/––/– – ––

This is to certify that the thesis entitled **The Study of Automatic Authentication of Printed Security Documents**, submitted by **Biswajit Halder** under Ph D Registration Number - Regn/Comp. Sc./Sc/484, the University of Burdwan, is a record of bona fide research work under my supervision and I consider it worthy of consideration for the award of the degree of Doctor of Philosophy of the Institute. This work has not been submitted earlier to any other Institute or this University for any degree or diploma.

**Dr. Utpal Garain**
Associate Professor
CVPR Unit
Indian Statistical Institute
Kolkata - 700 108, India

# CERTIFICATE

––/––/– – ––

This is to certify that the thesis entitled **The Study of Automatic Authentication of Printed Security Documents**, submitted by **Biswajit Halder** under Ph D Registration Number - Regn/Comp. Sc./Sc/484, the University of Burdwan, is a record of bona fide research work under my supervision and I consider it worthy of consideration for the award of the degree of Doctor of Philosophy of the Institute. This work has not been submitted earlier to any other Institute or this University for any degree or diploma.

**Dr. Abhoy Ch. Mondal**
Associate Professor
Dept. of Computer Science
University of Burdwan
Burdwan - 713 104, India

# DECLARATION

In accordance with the appropriate regulations, I, Mr. Biswajit Halder under Ph D Registration Number - Regn/Comp. Sc./Sc/484 submits my thesis and I declare that:

a. The work contained in the thesis is original and has been done by myself under the general supervision of my supervisor.

b. The work has not been submitted to any other Institute for any degree or diploma.

c. I have followed the guidelines provided by the Institute in writing the thesis.

d. I have conformed to the norms and guidelines given in the Ethical Code of Conduct of the Institute.

e. Whenever I have used materials (data, theoretical analysis, and text) from other sources, I have given due credit to them by citing them in the text of the thesis and giving their details in the references.

f. Whenever I have quoted written materials from other sources, I have put them under quotation marks and given due credit to the sources by citing them and giving required details in the references.

Biswajit Halder
Research Scholar
Department of Computer Science
The University of Burdwan
Burdwan - 713104

*Dedicated to My Parents and Family*

# ACKNOWLEDGMENT

I am deeply obliged to my external guide Dr. Utpal Garain for his dedication, encouragement, inspiration and supervision without which this thesis would not have been possible. I am greatly thankful to my internal guide Dr. Abhoy Ch. Mondal for his concern, encouragement and advice. My sincere thanks also forwarded to Prof. (Dr.) B.B. Choudhuri for granting me to study in Computer Vision and Pattern Recognition (CVPR) Unit, Indian Statistical Institute (ISI), Kolkata. I would also like to thank questioned document examiners of the Department of Forensic Sciences, Kolkata, India, who had made great contribution by sharing ideas, comments and materials. My dearest thanks go to Ankush Roy (University of Alberta, Canada) and Rajkumar Darbar (IIT, Kharagpur) for their invaluable help me. My thanks also go to Department of Information Technology, Durgapur Institute of Advanced Technology and Management (DIATM) for granting me required study leave for conducting my research. It gives me immense pleasure to thank the head of the Dept. of Computer Science, University of Burdwan, for extending me all the possible facilities to carry out the research work. I also extend my sincere thanks to all faculty members and administrative staffs of CVPR Unit, ISI Kolkata and Department of Computer Science, Burdwan University (BU). Finally, I forward special thanks to my family and all friends who cared, advised and couraged me Ph. D. study.

Biswajit Halder

# Abstract

Now-a-days, counterfeiting of security documents and old manuscripts are a serious problem in almost every country. The security documents, like currency notes, deeds, wills, bank-drafts, bank-checks, postal stamps, tickets, certificate etc., are valuable assets of our society. Therefore, authenticity of these documents is a big concern. Authentication technologies are becoming even more important for these print security documents, with the decline of copier, scanner, printer and the rise of other centralized printing devices which may be accessible to people without the required authority. Sophisticated authenticated technology can protect millions of value documents worldwide from forgery, tampering and counterfeiting. This study gives important security solution for automatic authentication which covert with different security features signification.

In the traditional paper-based world, when an authentic document is generated, it is usually signed, issued or approved by one or more authorized persons, with their signatures or seals to show the authenticity. Generally, the document with original signatures is considered to be original, authentic or legitimate. In the printed world, there are also requirements for such signatures to show the authenticity and originality of a document. The Questioned Document Examiner being asked to determine questions of authenticity in regards to not only questioned signatures but also to examine questioned typewriting, inks, writing instruments, paper, alterations of documents, erasures or other obliterations and also determine the relative date of a particular writing. Generally, Forensic expert or question documents examiners are investigating to analysis by different ways like as laboratory equipments, age verification, paper or ink component checking, handwriting and typewriting analysis and also check by forgery equipments. All these experiments are a manual basis or time consuming approach. Some time it's required quick decisions. So there is a need to design a system that is helpful to recognition these documents authentication by automatically. The systems are also demand for checking with fast speed, less time and also much economical. This study reports are tried to integrate a future scope to generate this type of system.

There are some excellent attempts in building robust automatic document authentication system in industry or academic. Few intelligent recognition systems are currently commercially available only for currency note only but these are not so good outcome. However, there is only limited research effort for recognition automatically for different

types of security document. In addition, diversity of archived different security documents and their security features poses many challenges to document recognition and analysis. Hence in this work, we explore novel approaches for automatic authentication of security documents that vary in features type and quality for printing and as well as paper.

We proposed here self adaptable authentication framework for recognition originality of security document. These processes follow on scan document images. Our study focus on printed documents which comprise security features printed, applied or otherwise provided on its. The authentication method is performed by acquiring features through image processing techniques. The region of interest is selected to encompass at least part of the features. In general, we apply linear or non linear learning procedures on captured features for classification in different group. This is achieved by Pattern Reorganization principles. A set of optimal features are trained for the classifier. Recognition results are also presented on synthetic as well as real life documents such that Indian currency note and old magazines to demonstrate the performance of the reorganization.

**Keywords:** Security Document, Authentication, Document Processing, Pattern Recognition, Currency note recognition.

# Contents

# List of Figures

# List of Tables

# List of Symbols and Abbreviations

## List of Abbreviations

| | |
|---|---|
| UV | ultraviolet |
| SVM | Support Vector Machines |
| ANN | Artificial Neural Networks |
| NN | Neural Network |
| LDA | Linear Discriminant Analysis |
| RBI | Reserve Bank of India |
| CMYK | cyan, magenta, yellow, and black ink |
| ATM | Automatic teller machines |
| DCT | Discrete cosine transform |
| PCA | Principal component analysis |
| SVD | Singular value decomposition |
| OVI | optically variable ink |
| RBF | Radial basis function |
| MLP | Multi-Layer Perceptrons |
| MSE | Mean squared errors |
| lpi | line per inch |

| | |
|---|---|
| IHT | inverse half-toning techniques |
| WCSS | Within-cluster sum of squares |
| SSIM | structural similarity index measure |
| PSNR | peak-signal-to-noise ratio |
| Half tone | HT |
| RBF-NN | Radial basis function neural net |
| ROI | Region of Interest |
| poly | Polynomial |
| GLCM | Gray level co-occurrence matrix |
| BPNN | Back propagation neural network |
| RMSE | Root-mean-square-error |
| ROC | Region of Co-occurrence |
| ACO | Ant colony optimization |
| NC | No of samples or nodes |
| NT | No of decade or stations |
| VSC | Visual Spectral Comparator |
| DFT | Discrete Fourier transform |
| DCT | Discrete cosine trans- form |
| ATR | Attenuated total reflectance |
| HSI | Hue Saturation Intensity |
| WInHD | Wavelet-based Inverse Halftoning |
| LUT | Look-Up Table |

# List of Symbols

| | |
|---|---|
| $\mu$ | Mean |
| $C_G$ | genuine samples |
| $C_D$ | duplicate samples |
| $\gamma$ | kernel parameters |
| $V$ | Support vectors |
| $T$ | Test samples |
| $\omega_j(x)$ | Transfer or Activation function |
| $\omega_j$ | Weights |
| $\alpha$ | Learning parameter of NN |
| $\beta$ | Momentum of NN |
| $T$ | Target |
| $O$ | Output |
| $FS$ | Fourier spectrum |
| $w_h$ | Synaptic weight of NN |
| $\sigma_h$ | variance of Gaussian function |
| $M_t$ | co-occurrence matrix |
| $\tau$ | Pheromone value of ant |
| $\eta$ | Initial heuristic value of ant |
| $CM$ | convolution matrix |
| G | Genuine bank note |
| D | Duplicate bank note |

# Chapter 1

# Introduction

## 1.1 Introduction

The invention of 'paper' has greatly contributed to the development of our civilization. The paper mainly uses as the substrate in printing processes can be considered to be a mature technology. Printing technology facilitated the communication-revolution which reached deep into human thought and social interaction. It also encouraged to pursuit the personal privacy. This orientation to privacy was part of an emphasis on individual rights and freedoms that print helped to develop. In today's business transactions, academic purpose, administrative processes and also govt. & non-govt. sector, printed-paper documents play an important role. Print documents also injected with standardization, verifiability for communication which comes from one source and is disseminated to many geographically dispersed receivers. Printed paper documents that are used to prove evidence of something are security documents. Like, currency note is a security document of wealth, mark sheet or certificates are security documents of course work accomplished and bank cheques are security documents of money transaction etc. Security documents have security values. These documents should be genuine and should communicate the correct information for the situation. A misrepresenting

document is a document that conveys the wrong information about an individual, but accepted anyway. A faked or counterfeited security document is one that has been carefully made by a fraudster to look like a genuine document; the complexity of making such document is measured by the counterfeit difficulty level. The automatic detection of fake document is really a complicated task and research in the authentication of printed security documents is growing rapidly in recent times because of its huge commercial potential.

In every country, forgery document was adept from the earliest times when writing paper documents were the medium of communication. Earlier, photographers with very little knowledge about this field were tried to extend their profession into document examination. But on the period of late 19th century, first time U.S. courts had liberalized their policy of evidence to allow handwriting analysis, photographic evidence and typewriting analysis [1] [2]. The private document examiners were termed to modern questioned document examination. Around first third of the 20th century, under the acceptance by the courts in U.S., the public examiner and governmental laboratories first developed and expanded their work to dominant in this field of work [3]. Technical manuals and other scientific reports on those areas began to publish around that time. In India, document examination was started through fingerprint signature checking which introduced more than a century ago. On the period of 1900's the British Government of India created the post of Government Handwriting Expert of Bengal for identifying the handwritings on the secret documents. And the services of this office were thrown open to criminal and civil court case. Later these works have involved lot of science and the services of this office were felt in many other cases like secret censorship, including the detection of invisible writings etc. After independence, Central Forensic Science Laboratory (CSFL) was established at Calcutta. At the starting time, this laboratory was organised into few basic disciplines viz. Question Document examination, Forensic Physics, Forensic Chemistry, Forensic Biology and Forensic Ballistics. Later for growing demand, most

of the states have established new forensic science laboratories for improving the study and application of Forensic examination.Now-a-day these depatments have important role for haldeling question document.

Today's one of the primary threat to every country is the unauthorized reproduction of printed security documents for unlawful purpose. This replication may take place via document scanners with computers, photo copiers, or illegal printing operations. It is easier than ever to forge documents because technology is becoming more modernize or powerful on a daily basis. Computers, and people who are proficient in using them, can be very dangerous to any business if proper steps to ensure document integrity are not employed. Desktop publishing software has created enhanced capabilities for the creation of documents that look quite similar to the real item. Current costs of scanning equipment, personal computers and laser printers have put this technology well within reach of criminals. They scan a document by these systems and can manipulate variable information creating output that can easily be used for fraud. In the past few years, color copier has grown over into a device that is accessible to almost anyone. And no particular skill is required to operate a copier and this makes it the smoothest way document fraud today. Unfortunately, untrained persons will not be able to tell a duplicate from an original. Obviously, the main target for such duplicity or fraud security concern will be for checks and other bank notes. However, there are a number of other documents that should also be protected from forgery. Some of these documents include draft, certificates, tickets, birth and death certificates, titles, coupons, registrations form, and bond papers. Therefore, design of an easy and quick method for authentication of security documents would be of enormous help for the communities that everyday deal with huge number of security documents.

## 1.2 Problem Definition

Every security documents have security values, the probability that the document is genuine and communicates the right information for the situation. A misrepresenting document conveys the wrong information about an individual, but is accepted anyway. A faked or counterfeited security document is one that has been carefully made by a fraudster to look like a genuine document. The complexity of making the document is measured by the counterfeit difficulty value. With the advent of cheap and sophisticated scanning and printing technologies, the security document frauds are on rise.With today's availability of low-cost scanning devices, high-quality printers, advanced color copy machines and other sophisticate devices, it gets much easier to manipulate printed documents. This has a serious threat to every society and the economics of any nation. So, authentication of printed documents is a task of increasing importance [4]. Authentication of security document incorporated with different security level features which integrate into security documents by artwork, ink, paper, printer etc. Individual features level authentication are requiring to judge for final decision and importancy measure.

### 1.2.1 Scope of work

The work mainly focused on four different sub-areas of printed security documents i.e. security artwork or design, security printing technique, security paper and ink-age. In this section, we discuss problem-definition and scope of this thesis work which is stated that individual sub-areas in sub-section 1.2.1.1, sub-section 1.2.1.2, sub-section 1.2.1.3 and sub-section 1.2.1.4 respectively. Finally, scope of currency note authentication with above features is discussed separately on sub-section 1.3.5.

#### 1.2.1.1 Security art-work

Security artwork or security design contains a comprehensive set of design tools to add copy-resistant graphical safeguards to certificates, coupons, titles, tickets, monetary or

legal documents, and other valuable documents. These graphical effects created from specific types of security design software [5] which provide a lot of graphical imaging features on the body, background, and borders of the security documents [6], [7]. These are virtually unattainable to duplicate. Generally two types of security features are integrated into art-work based security documents i.e. intrinsic and extrinsic. Intrinsic features [8] are integrated into the printout by the normal document generation process. But extrinsic features are added solely for the task of securing document. Many types of extrinsic features can be found in literature [9], [10]. Unlike most of the security features are commonly used from long period of times, some are complex and some are newly incorporate day to day. The copy-resistant graphical features which will generally protect a document are - micro printing, fine-line reliefs, micro-text, guilloche patterns, security back printing, padlock icon, Prismatic colored backgrounds, Lacey geometric patterns, Void pantographs, Bleed Through Numbering, Prismatic Printing, warning border, Prismatic Printing etc. [1], [11]. Out of these, security features are continuously being updated and developed to defend against alteration, erasure, toner removal, photocopying, and counterfeiting. Choosing the right security features for specific security document can save from counterfeiters. Though many art-work based security features are available, we discuss some common and useful security design based features and their security competence.

*Micro Printing* is done with very fine line-screen and resolution that provides by fine-line or micro-text. It is nearly visible and fills in when photocopied or scanned. Designers utilize a special type font and size (i.e. small) so that the microprint line image is clearly readable under magnification, but will become blurred and unreadable when copied or scanned. Some of common micro printing base features are used in security documents i.e. micro-words, micro-print signature line, Guilloche or Geometric Lathe Work etc. Micro Words are a series of small words prints across the sheet. Micro-printed words document are hard to replicate. Here entire documents are repeated by a single text-line

or a series of text-lines. These lines of text may be printed on the front or back of the sheet. For geometric Lathe Work or Guilloche method, an ornamental pattern of two or more interlaced curved bands that produce a unique circular design, commonly seen on banknotes, certificates. Micro-print line or text printed so small that it cannot be detected without magnification. It appears as a solid line in naked eye which cannot reproduce micro-printed line or text by photocopier. Padlock Icon is a universal symbol within the check printing industry which also follow micro-print line or text feature. It is used to identify and define the multiple security features used on a specific check. Security Back Printing commonly referred to as a security screened backer where a word is reversed out of a lightly printed screen. When copied or scanned, the screen and the reversed out word will drop out and not reproduce. When ordering checks with this feature, Graphic Dimensions will ensure backer with the Federal Reserve Regulation CC [12, 13]. Bleed through Numbering prints in black or red and will penetrate the paper to create an image of the number on the back of the sheet. This image cannot be lifted or altered. Authenticity of a document can easily be verified without the use of any special agents. This feature can be used for both MICR [14] and Arabic numbering. Padlock Icon is a universal symbol which checks by printing industry. It does define the multiple security features for identification.

Out of these, security features are continuously being updated and developed to defend against alteration, erasure, toner removal, photocopying, and counterfeiting. Choosing the right security features from huge number of available features for specific security document is difficult task for designers. Most of cases automatic authentications proof are not yet done.

#### 1.2.1.2 Security Printing

Security printing is one of the important fields of printing industry which try to prevent forgery, tampering, or counterfeiting. It deals with lots of security documents like deeds,

certificate, Currency notes, bank cheques etc. Generally choose of printer is depending on document cost-value, quantity and artwork. Other characteristics are line thickness, density and ink color. Previously, printers are using the more traditional offset and flexographic presses. Recently many of the modern printing techniques used to protect the high-value security documents which have become commercially available. Printing ink also contribute important role for their consideration.Often security features of inks can be combined to increase the security level and make printing of secure documents more cost-effective. Some useful security printing inks are - color shifting ink, thermo-cromotic ink, Penetrating Inks etc. Color Shifting Ink shifts from one color to another when viewed from different viewing angle. Suitable to protect high security documents like vouchers, tickets, security labels, certificates and others. Thermo-chromotic inks is able to change color appearances twice when exposed to temperatures beyond two different temperature change points. Generally three types of thermo-chromic inks are available tri thermo-chromic ink, thermo-chromic reversible and thermo-chromic irreversible. Tri-thermo-chromic ink is sophisticated appearing in three colors. Reversible thermo-chromic ink gives the effect "color to color to transparent" whereas irreversible thermo-chromic ink gives the effect "color to color to color". Penetrating Inks contain a special dye which penetrates into paper fibbers. This inking feature is used in intaglio printing process. So we try to incorporate some ink based features for verification of printing technique.

Some pattern follows for particular type of security-printing. Fine relief-line printing pattern, latent image, line HT etc. are commonly used. Fine relief-line printing pattern is a series of lines and curves which originate from a high resolution scanned image. These relief-line images are controlled through a visual dialog panel and it's typically discarded. Designer are controlled to specify it's line spacing, front and back angle, draw direction, line thickness, density, ink color etc. To give additional complexity, the relief line can be solid or screened, and reversed-out of another background color or image.

Interesting effects is coming by adjusting drape angle. This type of security document print through indirect printing process like indirect letterpress, dry offset. Photographic processes are applied to the printing plate, producing a relief surface, only the raised parts get into contact with the rubber blanket. Intaglio printing is one such printing technique which efficiently prints on important security document features like latent image. Latent image is such type of pattern which protects security documents against counterfeiting. Three-dimensional characteristic of latent images are illuminating document authenticity at acute angles of observation. Intaglio Printing creates results that rise above (referred to as relief) the surface of the unprinted paper. This physical relief can contribute to the security of the document. Photo copiers can reasonably simulate intaglio printing by the thickness of toner on the surface of the sheet but the reverse will not reflect the recessed area under the printing that is physical with Intaglio Printing. That is affected on intaglio printed type-face by its edge, smoothness, contrast etc. These effect can computationally extracted and judged by latent image, relief-line printing type-face etc. which are printed by intaglio or other printng process.

#### 1.2.1.3 Security Paper

Microscopic level, paper shows a fibrous texture of meshed plant fibres of varying lengths depending on the raw materials. Mainly, standard paper is made from wood pulp. But security paper is made from plant material such as reeds, flax and cotton plants. The fibres are stronger and visibly longer when it viewed under a microscope [15]. Security paper is also adding such type of special features during its manufacture. It's identified or authenticates a document as original. Through conventional reproduction process, security paper based print document cannot be reproduced or copied. Sometimes hidden security paper based features will show up during photocoping or scanning [16]. One of the oldest and quite common security features that is still very efficient is watermark. Visible fibers, Invisible Fibers, Security threads, holographic strips, iridescent coating,

biometric strips etc. are another type of paper based security features [16]. Designer select these security features on the basis of document cost, art-work and quantity. In this section, we discuss some common and important security features which embedded in the paper.

Watermarks are consisting of thick and thin areas in the paper that in transmission respectively appears darker and lighter than the surrounding paper. Watermarks reproduce a pattern in paper because of the variations of fibres density within the sheet. Sophisticated design makes the watermarks extremely difficult to counterfeit [17]. Security threads are an integral part of the paper. These are embedded in the paper when the sheet is formed, either completely or in a window. It can be customized by printing, demetallizing or a holographic effect. Thread width may vary from 1 to 6 mm. Covert security can be obtained by adding fluorescence [18]. Visible Fibres are added to paper which make difficult to duplication. These types of fibres are visible in ordinary light and can be of various colors, lengths and densities. Generally, fibre colors are different than the printed image. So, it can complicate counterfeiting by making it necessary to print with many extra colors. It is seen that color copiers can duplicate fibers based security documents but the created image will be on the document surface and easily detectable by touch. True fibres are imbedded into the sheet and are not bumps on the surface. Also, because fibers are added when the paper is being made and it generally are present on both sides of the document. Careless counterfeiters are likely to ignore the need to copy the back as well as the face. Using fibres as an overt (obvious) feature will signal the criminal that the document is protected and perhaps, deter fraudulent attempts. Invisible Fibres are invisible in ordinary light, but in ultra violet (black) light this covert feature is response. The fibres become quite visible when the document is placed under black light. Invisible fibres can be also added array of colors, lengths and densities. The inclusion of color makes simulation a much more difficult reproduce. These fibres are intended to restrict photo copy replication. These invisible fibres are a

very effective layer which completes a complex security system. So color level or physical level statistical measurements of these paper pulps or security thread can judge paper based authentication. Image processing and pattern reorganisation can help this process for machine level authentication.

#### 1.2.1.4 Ink-age

The ink age determination problem or existences of document quality are old question for document forensic. Particularly those documents were generating many decades ago or hundred years ago [19], [20]. Question document examiners are generally concerned about its originator and generation time. Answer of the first question may assist through signature verification or handwriting analysis [21], [22]. But answer of the second question is still unsolved. In literature [23], [19] it's seen that ink age determinant done through its chemical component analysis. Forensic experts try to ascertain the authenticity of these documents through several means [6], one of which is to check the relative or absolute age of a questioned document by determining the age of the printing ink thereon. Ink age determination has been a standard field of study in forensic science or question document examiners [11]. So far the forensic experts have been following different chemical techniques [1], [24] for determination of ink age. Where, the experiment quantify through ink and chemical component. These types of experiments are destructive in nature. This examination is not reasonably cooperative for those documents which require in future. Either if the organization deal with huge number of documents then it's very difficult task to solve. So machine label ink-age determination is required to solving.

#### 1.2.1.5 Currency-note authentication

Automatic paper currency recognition is important for modern banking system. It would require for many application like in Automated Teller Machines (ATM) and automatic

goods-seller machines. Normally, banknotes include maximum number of security features which can be included by art-work or design, sequence of printing technique, security paper or printing security-ink. On previous subsections we already discussed problem definition and scope of works about art-work, paper based authentication and printing techniques verification.Security ink, particularly currency note printing have a various types which can solve many document fraud and counterfeit issues. Often security features of inks can be combined to increase the security level and make printing of secure documents more cost-effective. Penetrating Inks is such type of security inks which contain special dyes that penetrates into paper fibbers. This feature is most commonly used in consecutive number applications. Photo-chromic Security ink is invisible under artificial light and will appear visible once exposed to normal sunlight or UV or black light. Once the ink is not exposed to sunlight or UV light, the ink will return to the transparent appearance for one or two minutes. This reversible effect is suitable for anti-counterfeiting of such type of documents. UV-fluorescent security inks are commonly used on these documents. Special type of effect will appear for this ink. When printed area are exposed by UV light then UV-fluorescent ink color are visible.

Initially, currency note features can be easily checked by common procedures like "Look, Feel and Tilt" [25], [26]. Also some different tests are based on the observation of the note with the naked eye in reflected or transmitted light and on the perception by touch on its surface. Banknotes also include other basic security features based on the response of different parts of the front and reverse sides to light from the infrared and ultraviolet spectra. In this literature, it's observed that some patents are obtainable for authentication of currency notes from that period. Because of the commercial interest, these patents express little technical and experimental details and this has restricted the research community to judge the performance of the systems. But, these are not sufficient now a day. Before long period of time counterfeiting currency note detection is an important field of research. Counterfeiting of currency note are increasing and rais-

ing complexity day to day wherein hi-tech instruments and technology are being used in printing of counterfeit notes. The challenges facing currency forensics are many, the effort required to detect counterfeit banknotes is directly correlated to the skill level of those producing the counterfeits, which shows no sign of stopping.

## 1.3 Previous work

For checking authentication of any security document in question, Question document examiners have so far relied upon the human experts. Forensic departments in different countries have opened up special branch for this purpose. They in general involve different devices (for example, UV lamp, magnifying glass, IR detector, etc.) [27] and observe certain properties of the security document in question. They arrange these observations to come up with a decision. However, involvement of such authentication process is quite difficult but not impossible where a department/office daily has to deal with a huge number of such documents. Moreover, involvement of forensic people requires lots of administrative and legal steps like filing a legal case on finding a document in question, sending the document in original to the police people, waiting for expert's view, etc. making the entire process a lengthy one. Every day, they also deal huge number of security documents. Therefore, design of an easy and quick method for authentication of security documents would be of enormous help for such communities [28, 29]. So, in this section we discuss about related work on automatic authentic technique. Here, we particularly follow separate subsection for authentication of artwork, printer and paper. Related work of ink-age determination and currency note authentication is also mentioned in separate sub-section.

### 1.3.1 Authentication for security artwork

Nature of security-design based features varies from one class of document to others class of document. A counterfeiter's primary objective is to make possible reproductions, by

mastering known security features. Authentication of art-work based security features are classified of three different categories: first one, immediately detectable through human senses, second one, hidden from normal view of our senses, detectable using basic tools, such as a magnifying glass, microscope etc. and third one, intrinsic characteristics resulting from the manufacturing process and the interaction of raw materials. The first level security is directed towards the human senses, typically sight and touch, but also sound, it is noteworthy that the majority of security components are found in this level. Presently, most of the authentication process relied upon the human expertise knowledge. Forensic departments in different countries have opened up special branch for this purpose. The second level security features are approximately hidden from view of our senses without using basic equipment. In inspection phase generally involves different devices (for example, UV lamp, magnifying glass, IR detector, etc.) [27] and observe certain properties of the document in question. Then it arrange these observations to come up with a decision. There is a vast variety of features existing to make security documents relatively protected even through first or second line inspection, e.g. using special sort of paper, eventually integrating a watermark, by adding holographic images [30] specialized printing techniques [28] and other physical and chemical signatures [31]. Many other types of features maintain in literature [9,10,32]. The third level characteristics are reserved for Question document examiners (QDE) or forensic examiners, such measures often are inherent characteristics resulting from the raw materials and printing processes used. The art-work based features depend upon the importance of a document. For example, currency notes are very high level of security aspects whereas bank cheques or drafts fall in the middle of this complicacy scale. Lottery tickets, postal stamps, etc. are somewhat less complicated. Authentications of currency note are importance and signification value for long period of times. It is seen that maximum numbers of art-work based features are entrenched in currency note which broadly followed by other security documents. Research on manufacturing secure currency notes is in-deed an old research

topic. Because of the commercial nature for this research studies were patented rather than published in scientific journals [30]. A department/office who is dealing with a huge number of security documents, examination of such process is quite difficult but not impossible to judge. So, huge demand of this study is to generate automatic system. Present scenario, the research work in this aspect focuses on automatic machine authentication of printed security documents. The entire experiment is carried out by image processing and pattern recognition technique. A trend is observed that an image is acquired, it is pre-processed then classified by pattern recognition techniques and finally the output result is obtained. Various methods are employed for these experiments. Researchers try to incorporate maximum number of art-work or design based security features for this authentication. The impotency and correct combination to use is subjective to the various security documents in question. In literatures it generally seen that different workflow is employed on digital currency forensics system to classify the bank-note. Many art-work features are involved here. Most of cases, image processing techniques are employed to extract key characteristics features i.e. numerical measures which computationally describing a banknote image. Color-level features, shape and texture descriptor are common areas which engross for this purpose. Color measurements describe the level of color by creation of intensity or grey-level histogram, shape descriptors are key color characteristics. Texture describes the pattern of pixel colour and their relationship with one another. It is found that using a vector apace composed of both texture and colour features may improve classification accuracy [33]. Modified canny edge detection is used [34] for removal of noise and background from captured images. The Canny algorithm marks a point as an edge if the amplitude is larger than its neighbors without checking that the differences are higher than expected. This results in the algorithm being sensitive to weak edges such as where the colors are similar or where there is blurring in the picture. Beusekom et al very recently reported [29], forgery detection using text-line information is presented. Here, authors are using text-line ro-

tation and alignment to detect documents that have been changed with a malicious intent. By measuring and detecting mis-rotations and mis-alignments of two text-line features, an automated approach for verification of documents is proposed in this paper. Another paper of same authors [29] authentication of documents can be done by detecting the printing device used to generate the print-out. Here authors pointed out that many manufacturers of color laser printers and copiers designed their devices in a way to integrate a unique tracking pattern in each print-out. This pattern is used to identify the exact device where the print-out originates from. Authors were also comparing two base patterns from two different print-outs to verify if two print-outs come from the same printer. They mention automatic decoding of the base pattern to extract the serial number which was maintained at the time and date when the document was printed. Signification evaluation accuracies result for detecting the printer class reported here.

### 1.3.2 Authentication of printer

With the recent use of digital imaging techniques for the forensic examination of documents, determination of underlying printing technology of a security document has gained significant attention of the research community. Two different approaches are prevalent for identifying printing technology. Often, the first step is document forensics which focuses on source identification and second step is document forgery detection. First level of approach makes use of gray level or colour of image pixels for discriminating a specific printing technique from others. The aim is to trace the source of the documents which mainly include as printer identification. For example, Mikkilineni et al [35–37] considered small printed areas of a document and print quality defects are modelled as texture. Texture features are then used to classify the model and manufacturer of the printed document used to print a document in question. Somewhere statistical individuality of the printed-characters is used to recognize the source brands and models. Like, Mikkilineni et al. ( 2010) used gray-level co-occurrence matrices (GLCM) features, vari-

ance/entropy, and discrete Fourier transform (DFT) [38] features to identify the printer source. Both the texture and edge based gray-level features [4] were used to compute the feature vectors. Classification was done by using three different classifiers namely, decision tree, multi-layer perceptron, and support vector machines. Experiment nicely illustrated the effect of scanning resolution, the kind of features used and the particular classification approach on the accuracy of identifying printers. Here, authors suggested that apart from gray-level features use of color properties of documents may help to achieve better accuracy. Use of color features to a limited extent is available in the study conducted by Dasari and Chakravarthy [39]. They used HSV color space and, in particular, hue images at high-resolution to distinguish between the different printing processes. Banding frequency is another characteristic used to identify the source as a laser printer [40]. Banding, which appears as non-uniform light and dark lines perpendicular to the printing direction is caused by fluctuations in the rotating angular velocity of the organic photo conductor (OPC) drum and by errors in the gear transmission mechanism. Schreyer et al. [41–43] also worked on detecting the printing technique where they classified and also analyzed between printed documents and copied versions of printed documents. Here, authors were using discrete cosine trans- form (DCT) features, and shown good performance even scanning with moderate resolutions of 400 dpi. Kee and Farid [44] attempt to printer-model for characters to detect documents forged by different geometric degradations. Different geometric degradations generated by different printing processes. This method is heavily dependent on availability of character images in the document in question. They used principal component analysis (PCA) and singular value decomposition (SVD) to model the degradation of a page caused by printing, and the resulting printer profile was then used to distinguish between characters generated from different printers. However, it was also experienced that as a printer profile depends on its toner level, different profiles corresponding to different toner levels are to be made for the same printer. Geometric/structural information is also used by

Li et al [45,46] for identifying colour printers. They found that the dotted motifs can be used to characterize printers of different makes. The authors even observed that printers of different serial numbers of the same make result in distinctive dotted patterns. Tweedy [47] investigated the use of such a coded pattern on each color laser copy so that the forensic people easily and uniquely identify the make, model, and serial number of the copier used.

On second step i.e. document forgery could involve changing, adding, or deleting some information on the document or replacing an entire page with a counterfeited page [48]. Kong et al. [49] established a projective transformation model to estimate such geometric distortions. Bulan et al. [50] computed the geometric distortion signature from printed halftone images to trace the source laser printer by estimating the variations in the center positions of halftone dots. For pasting and reprinting forgery operations, character location distortion is often introduced. Beusekon et al. [51] presented a technique for extracting text lines and alignment lines for document inspection. Another frequent intrinsic document feature is handwriting which contain as notices or signatures on many printed documents. Two related common questions rose: off-line writer identification and off-line signature verification. This problem consists in identifying the writer of a document in question. Schomaker et al. [52–54] specified this problem and proposed solution by using a previously trained model of the writer's handwriting. In this context, on-line means that only the image of the signature or the handwriting is available, in contrast to online data, where stroke information is also used. In the literature [55,56] of signature verification the question is whether a signature on a document has been generated by the person claimed by the signature or if someone else forged the signature. A machine-vision based print quality analyzes by Oliver and Chen [57] whose derive a particular printer's characteristics and identify printing technology. Seven different digital (ink jet and dry/liquid xerography) and impact (computer-to-plate offset lithography) printing technologies were considered and their characteristics were quantitatively ana-

lyzed using features like line width/raggedness and over-spray; dot roundness/perimeter and number of satellite drops, image sharpness and image growth (positive versus negative prints). Experiment showed that it was possible to resolve a unique print quality signature which enables differentiation of one printer technology/supplier from another. Though no such framework was proposed towards how machine can take decision about the printing technology but these studies demonstrated the potential of a machine-vision based approach in the context of digital printing forensic document examination.

### 1.3.3 Authentication of Security paper

The previous efforts attempted to exploit several security features but the paper material itself was hardly consulted for authenticating the document in question. Earlier, only some experiments have done through micro-scope, UV lamp Test. Paper, at the microscopic level shows a fibrous texture of meshed plant fibres of varying lengths depending on the raw materials. Standard paper is made predominantly from wood pulp whereas security paper is made from plant material such as reeds, flax and cotton plants. So, the forensic experts often check the intended paper quality by physical contacts and sometimes, though manually, they check the illumination, brightness and density of the paper pulp in order to authentic the paper of the document in question. Somewhere forensic observation made through microscope that the fibres are stronger, and visibly longer. [15]. During WWII, the Nazi regimes had succeeded in creating counterfeits British pound which was passable by lamp test but, countless experiments discovered their ingredient, later [58]. So this was very old question for paper level authentication. But, the research in machine aided authentication of security paper documents is still limited.

Some patents [59–61] are available for paper based authentication of currency notes. But these patents or available systems do not reveal any technical details or experimental results for commercial interests. Though, it has limited the advancement of the research

in this field but, some research labs have recently started studying this problem [62]. For any security paper documents including the banknotes, the paper itself plays a crucial role in proving some kind of security document [18]. The paper based security is normally achieved by embedding certain special ingredients to the paper material during its manufacturing process. Previously many researchers worked on the embedded water-mark and security thread that are hard to duplicate and filed patents on how to authenticate the watermark and or security thread [59, 60] and [61]. Colour optical pulp (or fibre) which embedded in the paper defines a certain kind of characteristics of the paper. Security fibres may be metallic or photo-chromic which are luminescent under ultraviolet (UV) ray and therefore, visible when the paper is scanned under UV or illuminated light ray. More elaborate review on security papers can be found in [63]. A two photon microscope is used, an excitement method used to view photons emitted [18]. Their findings show that fluorescence aspects of genuine banknotes differ quite considerably in wavelength and amount of photons emitted than fake paper. In recent time, Takalo et al. [64] is reported a novel method for distinguishing counterfeit banknotes from genuine samples which is based on analyzing differences in the networks of paper fibbers. Here, authors proposed curve let-based algorithm for measuring the overall distribution of fibber orientation and quantifying its anisotropy. This technique makes possible to distinguish forgeries from genuine samples. Author are also investigated the watermarks techniques by comparing genuine and counterfeit $\varepsilon$ 50 banknotes.

### 1.3.4 Ink-age determination

In recent time Neumann et al. [65] tried to estimation age by ink component. They followed that the ink ingredients typically differ from one currency to the next, ink needs to be both durable and difficult to copy. This requires top secret unique blends of raw materials, subsequently, each combination emits different levels of radiation. The US Secret Service over the last century has been building an ink library for forensic

purposes [65] this library supports analysis by providing a reference point allowing for both discrimination of ink compositions, and estimation of age. Others recently proposed methods can be found in [66]. The essence of these techniques is to measure or observe several chemical properties or changes in properties upon use of different chemicals and based on these measures or observations a decision is taken. For case-based experiment these techniques have proved their reliability to an extent. However, many of these methods are destructive in nature and not suitable especially when the document in question has to be preserved as it is. Methods are not fully automated and therefore, they require human intervention.

### 1.3.5 Currency-Note Authentication

On the contrary, the researchers are also reporting details of their methods and experiments which have so far dealt with recognition currency of different countries (U.S. dollars, Euro notes, etc.), and different denominations for a given currency. There are a few studies which address this problem of recognition and give details about their methods and experiments. Vila et al [25] proposed a semi-automatic approach for characterizing and distinguishing original and fake Euro notes which is based on the analysis of several areas of the banknotes using a Fourier transformed infra-red spectrometer with a microscopic ATR (Attenuated total reflectance) objective. They considered four different regions of a note and observed that fake notes are easily identifiable from the analysis of the spectra corresponding to the four regions. Later on, the authors in [67] describe another system for authenticating Bangladeshi Bank Notes. They assume that original currencies under test have the bank name printed in micro letter print. They scan this part (the region where the bank name should be) using a grid scanner and the textual images are fed in an optical character recognition engine that matches characters with prototypes. Proposed algorithm is mainly dependent on that feature which makes the system very sensitive.Somewhere security printing inks have played certain role for

such authentication. For example optically variable ink (OVI), staining ink, pearl lustre ink etc are definite role for different bank note printing [68] like French 50 frank note, Swiss 50 frank note, Dutch 100 guilder note. But raman spectroscopy failed to determine the difference between fake and genuine Euros inks as shown in [68] because similar colour spectra obtained for the fake and genuine euro banknote but one thing of Raman Spectra still utilised for fake banknote i.e. different areas of fake Euro banknotes for the same colour show similar spectra. For same purpose, Neural Networks [4], [59,69], [70,71] and [72], Genetic Algorithm [73] and Hidden Markov Model [74] have so far been used. These studies have successfully addressed the currency or denomination recognition problem but did not consider whether the input bank note is genuine or fake. There are a few studies which address this problem of recognition and give details about their methods and experiments.

In literature, it is seen that some patents are available for authentication of currency notes. U.S. Patent issued in 1857 [69] may be the earliest attempt of an optical method of manufacturing secure paper money. It involved using paper tinted to absorb light, and printing ink that also absorbed light rather than reflecting it so that clear photographic copies could not be made. Many researchers assumed that the embedded water-mark and security thread are hard to duplicate and filed patents on how to authenticate the watermark and or security thread [59], [60] and [61]. The patent in [75] described an approach that is based on the reflective property of the bank notes in question. A series of light source is placed that provides wavelength of varying illumination, which is used to measure the reflective and refractive response of the currency note images. This data is then compared with the pre-calculated data on original and fake notes to report the authenticity of the notes. On the other hand, the patent in [76] proposed a method to verify US currency notes by analysing different aspects of the ink used for printing. But these patents or available systems do not reveal any technical details or experimental results for commercial interests. These patents also express little technical and experi-

mental details and this has restricted the research community to judge the performance of the systems.

In recent time, some commercial system [77] is available in the market, E-Brochure one of them which are using to detect fake notes. In order to authenticate a paper currency, the system makes use of several devices like magnetic level circuits, visual light sensors, light intensity compensation circuits, Infra-Red ray sensors, Ultraviolet Ray Sensors and many other technical light sensors. These hardware components function simultaneously and compare crucial security features like ones from the based materials like ink, paper, security thread, fluorescent resins, secret watermarks, micro prints, intaglio patterns, etc., against the data stored in standard memory bank which keeps information about the characteristics of genuine and counterfeit notes. For a given note, if the features match to those of the real ones, the system passes the note as genuine otherwise, the machine saves the features match those of counterfeits and these notes are rejected.

## 1.4 Contribution of this Thesis

Though several efforts have previously been reported in the literature on security document authentication, focused research on automatic authentication technique as reported in this thesis could be considered as a novel study. Most of the earlier techniques of such examination are quite difficult while dealing with huge numbers of such documents. Involvement of forensic people often requires several steps like filing a legal case on finding a document in question, sending the document in original to the police, waiting for an expert's view, etc., thus making the entire process a lengthy one. In present scenario, the automatic document processing is highly required for forgery detection. Hence, automatic checking or authentication machine is required for question document examination.

The main contribution of our study is to focus on maximum number of security features of various type security documents for automatic authentication. We developed different

automatic authentication frameworks which will detect the forgery document. Proposed work has been incorporated on different sections like -

*security design verification:* Earlier several studies have been reported on automatic processing of security documents but authentication of microprint line or line HT art-work based security documents has never been attempted. This portion presents a pioneering effort to involve machine for authentication of security documents and line HT document detection. Criminals' efforts for generating fraudulent version of such security documents are on the rise. A particular class of documents i.e. microprint line based security documents has been considered for first experiment. IHT has been proposed for line HT documents verification for second experiments. Both studies attempt to develop a general framework for automatic verification of authenticity of such types of security documents.

*Printer detection:* Printing of security document provides important checkup for authentication of the security documents. Though, the printing technique that is hard to replicate because some of its inherent characteristics. But it's not imposable with the help of update and modern technology. Forensic experts often take their decision by checking these effects of printing on currency note with mostly the help of a microscope. Automatic verification is required for modern era. So our proposal is pointing towards this direction. In this portion of study, the entire approach is based on examining the characteristic a particular printer by consider three different aspects like geometric properties, gray-level features, and color properties. Most importantly, we attempt to closely follow the practice of the question document examiners in detecting printers and try to simulate the same in a machine-vision based framework. Most of the aspects that the forensic experts look for in order to identify the expected printing technique are computationally grabbed and a system is configured to give the decision about the authentication of the printing process.This decision is done by linear or non-linear classifier. This non-linear classification is done by using Support Vector Machines and

Neural Nets. The discriminatory power of the selected features in authenticating the printing process is tested using the Linear Discriminate Analysis. Experimental results show that the proposed framework provides a highly accurate framework for authenticating the printing process in bank notes

*Paper base detection:* Another experiment is paper based authentication checking where we considered again real bank notes sample. Here, we have considered fluorescent paper pulps which are visible in the UV scanned document image. These paper pulps play crucial role in authenticating the paper. In a counterfeit note, if the paper is very deferent from the genuine one, these pulps may not be seen at all. In a high quality counterfeiting, these pulps came as very bright spots and their shapes show significant deference with respect to the pulp marks of the genuine. Therefore, the illumination and shape of these paper pulps are important in characterizing a note paper as genuine or fake.Our study is incorporated with these signification characteristic of paper pulps.

*Ink age determination:* Experiment on printing ink can also determinate its age. Though, in previous works, the experiments were based on ball pen ink or handwritten based manuscript, but here, we have experimented on printing ink. In our proposed method we have basically analysised of ink colors and noted the changes of different colors with age. For example, a considerable old document become older much more yellowish and brownish than a fresh one. The color shade and brightness are affected on a document with its age and this in turn affects hue and saturation level of the ink colors. However, instead of doing all these analysis by chemically and manually, we attempt to concentrate on statistical analysis of relevant color features that are computationally captured from scanned images. Apart from gray level analysis, we consider two other color spaces namely, RGB and HSV. The work attempted to give a decision on its age and also give guideline of changing effect. We have also tried to incorporate a mathematical model for the color base changing effect on documents. Consecutive old editions of LIFE magazines cover page have been considered as a reference for this study.

*Currency Note Authentication:* Currency note Authentication is one example of the highest level of forgery detection. In recent years, we have seen a large increase of counterfeit currency notes. The process of detection of such notes is hideous and quite cumbersome. Initially, counting and sorting machine was introduced for automatic currency note detection. But, it has been seen to have certain technical limitations regarding checking procedure, sensing images and so on [63, 78]. As the speed of the operative system is high, it is difficult to properly detect the image of passing note. Because of highly heterogeneous concepts involved with Indian currencies, authentication of these notes poses a big technical challenge. In this portion of study, Indian bank notes are taken as reference to show how a system can be developed for discriminating fake notes from genuine ones. Image processing and pattern recognition techniques are used to design the overall approach. The ability of the embedded security aspects is thoroughly analysed for detecting fake currencies. Real forensic samples are involved in the experiment that shows a high precision machine can be developed for authentication of paper money. The system performance is reported in terms of both accuracy and processing speed. Comparison with human subjects namely forensic experts and bank employees clearly shows its applicability for mass checking of currency notes in the real world. The analysis of security features to protect counterfeiting highlights some facts that should be taken care of in future designing of currency notes.

The challenges faced in examining question document are so many, the effort required to detect forgery documents are directly correlated to the skill level of those producing the counterfeits, which shows no sign of stopping. This make the task of currency note detection very tough. Obviously, detection is not a trivial task, to forensically determine authenticity; one must possess a deep understanding of all features. To minimize and control the circulation of counterfeit banknotes, Automatic Teller Machines (ATM) and central banknote sorting machines incorporate banknote recognition software. Question document examiner faces lots of problem dealing with huge number of such documents

within specific time. In the present scenario, the research work in this aspect focuses on automatic machine authentication of printed security documents which are carried out by image processing and pattern recognition technique. For authentication of security documents, image processing techniques are used for feature-values extraction and these values are intregated by pattern recognition techniques for decision making. A combination of image processing, machine learning, and pattern recognition, the software is trained to check specific security features using learned threshold values. The accuracy rate and results analysis are to be made with elaborately discussion. Our study focuses towards this direction. We believe that this thesis will give more powerful decision making judgments and inner level analyses than those available online so far.

## 1.5 Organisation of this Thesis

This thesis is organized in eight chapters. The first chapter provides an overview of the general background and the problem setting. The previous work and the major contributions are also briefly described in this chapter. Chapter 2 we present overall proposed framework for automatic authentication of printed security documents. Chapter 3 describes artwork or microprint line based security document authentication. Printing technique detection describes in chapter 4. In chapter 5 we discuss our proposed method for security paper based authentication. Ink-age determination and model based approach presented in chapter 6. In chapter 7 one document-authentication based case study elaborately explains by real Indian currency note samples. Finally, Chapter 8 provides conclusion of the work, concluding results and scope for future work.

# Chapter 2

# A Framework for Machine Authentication of Security Documents

## 2.1 Introduction

After reviewing the literature in the previous chapter, it is found that there have been few studies reporting technical and experimental details on how to automatically authenticate security documents. Many studies rely on few features that also can be duplicated using today's high end technology. As for example, there are so many security features in currency notes, analysing only a certain aspect of the note security may not be a good choice. A complete integrated framework has been missing that looks into many aspects like security features in printing, ink, background artwork, ink-age, security thread, etc. Another general shortcoming in the existing studies is the use of synthetic data. Many authors generate sample at the lab to test their algorithms. Therefore, performances of these algorithms on real samples are not yet known.

All these shortcomings motivate us to take up the present research. This research-work

attempts to formulate a different framework for security documents authentication in different sub-areas i.e. art-work, printing, paper, ink-age etc. The entire approach is based on image processing and pattern recognition techniques.

Here, this study have generally considered three different aspects for features selection like geometric or physical-level properties, gray-level features and color properties for characterizing counterfeit documents. Most importantly, we have attempted to closely follow the practice of the question document examiners in detecting counterfeit security document and try to simulate the same in a machine-vision based framework. Most of the aspects that the forensic experts look for identifying the expected technique are computationally grabbed and a machine is configured to give the decision about the authentication of different level security documents or solved different quires of question document examiners. Proposed framework is also used for a practical problem namely, identification of fake document based on authentication of printing technique or currency note detection. The various experiments involve synthetic as well as real samples for genuine and fake documents. Results are computed and analyzed to bring out the potential of the proposed framework. In this chapter we discuss about the Existing Practices, experimental framework where we talk about on features selection and the general method of authentication.

## 2.2 Existing Practices

Security documents refer to documents containing incorporated security features within the document to protect the value of the document. Many identity documents contain security features such as passports, identity cards, and driving licences. Other examples of security documents include currency, bank cheques, social security cards, travel visas and lottery tickets, etc. The security features and techniques offering protection against authentication threats have been subdivided into basic and additional security features which states are encouraged to select items for providing an enhanced level of security.

*Basic security features:* Printers can offer customers five basic security features. The low-cost features include warning bands, padlock icons, microprint lines, security screens, and pantographs. Warning bands are stripes composed of reversed-out type around the perimeter of a document that point out what security features have been used on the page. A padlock icon serves the same purpose as warning bands. A microprint line is usually a line of one-point type. The content of a microprint line cannot be distinguished without a magnifier. *Additional security features* are not only graphics-related. These are many paper and ink combinations that make documents more difficult to forge. The paper used for a job can have embedded fluorescent fibbers that can only be seen under an ultraviolet (UV) light. Chemically reactive paper, specialty inks, and foil embossing are other examples of special features to ensure more safety. Most of the security documents include basic security features, so inspection or examination of these documents or features would require broadly.

*Security Document inspection* is an important aspect for security documents authentication. Protection of documents with security features depends on the relationship between document and inspection. Inspection of documents and valuable products for authenticity can be divided into three different phases: *1. Basic inspection phase* - The inspection of the document or product with the human senses only without additional equipment. First line inspection is aimed at unveiling counterfeits and forgeries (alterations) using public security features like portraits picture, security tread, watermarks, tactile intaglio printing, etcetera *2. Advance inspection phase* - The inspection of the document or product with the means of additional tools like a magnifier, an ultra violet source, a bar code reader, etc. In these cases the inspection requires a human inspector to judge the results. In the case of automatic teller machines (ATMs) and such like equipment, the inspection is automates but still have questions on their quality, speed and ability. *3.Specialist inspection phase* - The inspection of the document or product in laboratory conditions, using advanced knows how, sophisticated means (spectrometers,

microscopes, infrared radiation, etc.) and dedicated inspection facilities. But, Question document examiner of forensic department checks these documents on five parts - (i) examines the security feature of document design (ii) examines the security feature of printing process or ink pigments (iii) examines the security feature of paper (iv) examines document dating and (v) others (like physical dimension etc.).
Examination of security features are the challenges of these works. Sometimes require a quick judgment is requested. Mainly, these inspections are limited to forensic laboratories and only of practical only when suspected documents have been seized. However, execution of such a process is quite difficult while dealing with huge number of such documents. Involvement of forensic people often requires several steps like filing a legal case on finding a document in question, sending the document in original to the police, waiting for expert's view, etc. , thus making the entire process a lengthy one. In the present scenario, the automatic document processing is highly required for forgery detection. Our study ventures to go this direction.

## 2.3 Our Framework

The main outcome of this thesis is to apply the new proposed framework for machine authentication of printed security documents. [as depicted on figure 2.1]. The entire experiment is carried out by image processing and pattern recognition techniques. To the best of our knowledge, no effort towards this direction has been made so far. Here, we have taken references of the documents like bank cheque, currency notes, and old magazine cover page. In this section, we summarize the list of these problems considered in this thesis, and the way of solution obtained for those problems. At first, we study on microprint line based security features. Here, bank cheques take reference which attempts to provide an automatic method for authentication. The approach is based on pattern recognition principles by which relevant features are initially extracted from cheques and then using these features an algorithm for discrimination between genuine


**Figure 2.1:** Printed security document

and duplicate cheques is outlined. Security features are extracted from the scan bank-cheque color images. Both kinds of genuine and duplicate cheques are used to generate feature vectors. Distribution of these feature vectors is studied in the feature space and suitable classifier is designed. Experiment is conducted to confirm that the security features captured computationally are sufficiently robust to discriminate authentic vs. Duplicate document. This next experiment in this section is aimed to developing an inverse half toning technique (IHT) for authenticating line HT images. The method attempts to formulate a statistical measure in order to judge the quality of the image in question against the original image. We have considered line halftone image at different resolutions namely, line per inch (lpi) which are commonly used in practice.

We also attempt to formulate a general framework for authentication of the printing techniques in banknotes. The entire approach is based on examining the characteristic a particular printer by consider three different aspects like geometric properties, gray-level features, and color properties. Most importantly, we attempt to closely follow the practice of the questioned document examiners in detecting printers and try to simulate the same in a machine-vision based framework. Most of the aspects that the forensic experts look for identifying the expected printing technique are computationally grabbed and a system is configured to give the decision about the authentication of the printing

process. This framework is then used for a practical problem namely, identification of fake currency note based on authentication of printing technique. The experiment involves real samples of genuine and fake bank-notes. In this chapter, we discuss about the computation of features, implementation of the method, experimental results and analysis to bring out the potential of this proposed framework.

We also study to develop a machine assisted tool for authenticating the paper of a security document. The paper itself forms an important security feature for many security paper documents. In this study, paper pulps which are visible in the UV scanned image of the document play a crucial role in characterizing a paper material. Therefore, the pulps are first identified in the UV scanned image and this identification is done by borrowing ideas from rice grain detection method. Image processing and pattern recognition principles form the basis of this automatic method. Once the pulps are detected, shape and color features are extracted from them. Paper pulps coming from fake documents are significantly different from those of genuine documents in their shapes and colors. Using the shape and color features, a multilayer back propagation neural network is used to discriminate paper pulps as genuine or fake. The framework is tested with Indian banknote samples. Experiment shows that consideration of paper pulps is one of the crucial tests for authenticating paper money.

This study attempts to develop a general framework that makes use of image processing and pattern recognition principles for ink age determination in printed documents. The method is basically analysis of ink colors. The changes of different properties of colors with age are noted. For example, older a document is, much more yellowish and brownish it is. Dark noise also increases with age. The color shade and brightness are affected on a document with its age and this in turn affects hue and saturation level of the ink colors. The approach, at first, computationally extracts a set of suitable color feature and then analyzes them to properly associate them with ink age. Finally, a neural net is designed and trained to determine ages of unknown samples. The experiment has

done by LIFE magazines cover pages which published in five different decades. Test results show that a viable framework for involving machines in assisting human experts for determining age of printed documents.

One case study on automatically authenticate currency notes collectively incorporated with several security aspects as ink, security thread and an art work based feature set which have discussed on previous chapter. A large dataset has been used to test the system performance and a thorough analysis is provided by comparison with real forensic document examiners to show the applicability of our research in the real world scenario. Analysis is also done to report the speed of the system. Popular pattern recognition tools like the k-means, neural network and support vector machines are used to demonstrate the system performance. We have also shown the robustness of each feature in tackling the problem and thus pointing out the sensitivity of the features. The feature level analysis brings out important views that the note designing agencies could keep in mind regarding the susceptibility of the features to be duplicated. A sequential ordering of security features is also suggested that should be maintained while testing to maximize the performance accuracy. Involvement of real forensic samples is another significant aspect of this study. Indian bank notes are taken as a reference. In fact, because of highly heterogeneous concepts involved with Indian currencies, authentication of these notes poses a big technical challenge.

Next two subsections we will discuss on feature selection and general method of authentication respectively. On subsection 2.3.1 we talk about considering features level domain for security document authentication. We will discuss different propose-method for authentication proof on different sub-areas under subsection 2.3.2.

### 2.3.1 Feature selection

While preparing any security document, the designers embed certain features that are considered as security features. It is generally assumed that these features are difficult

to replicate or copy. Duplicity of a document is identified by checking these security features. Depending upon the importance of a document, nature of security features varies from one class of document to others. For instance, currency notes, legal deeds are having very high level of security aspects and security features in documents like lottery tickets, postal stamps, etc. are somewhat less complicated. Certain features in bank cheques or drafts are fall in the middle of this complicacy scale. Security features in security print documents are mostly incorporated in three distinct areas.

*Security design or background artwork:* Generally light fine-line printing or others security patterns are appear on background of security documents.[as depicted on figure 2.2] This type of printing is difficult to reproduce on scanning equipment or replicate by other printing methods. The use of an intricate fine-line/patterns background design (for example, Guilloche module [79]) is an essential part of any security document. Apart from using fine-line artwork, micro-sized typeset characters can also be included in the design. Occurrences of such characters are generally repeated throughout the background following certain periodic fashion (e.g. on straight lines). Deformation in shape characteristics of such micro-printed characters is often visible when they are scanned or copied. Sometimes special security icons or descriptive markers are used so that verification of cheques becomes an easy task for the bank people. For security design based authentication technique we concentrated on this artwork. Gray and color level features were importance contribution artwork based recognition. Another security artwork verification experiment is that to judge fine line half tone image. Most of these patterns are produced from continuous tone image. The design details of such patterns are not clear with the naked eye but become clear with magnification. Characteristics of these line patterns are line thickness, line density and ink colour. Fine-line design features are changed in the event of a photocopying attack. Fine line deformations of HT-image and texture pattern are also considered for physical level features which are mainly contribute on proposed inverse half-toning techniques (IHT).

*Printing Technique:* Different printing processes like offset, laser, intaglio etc give dif-

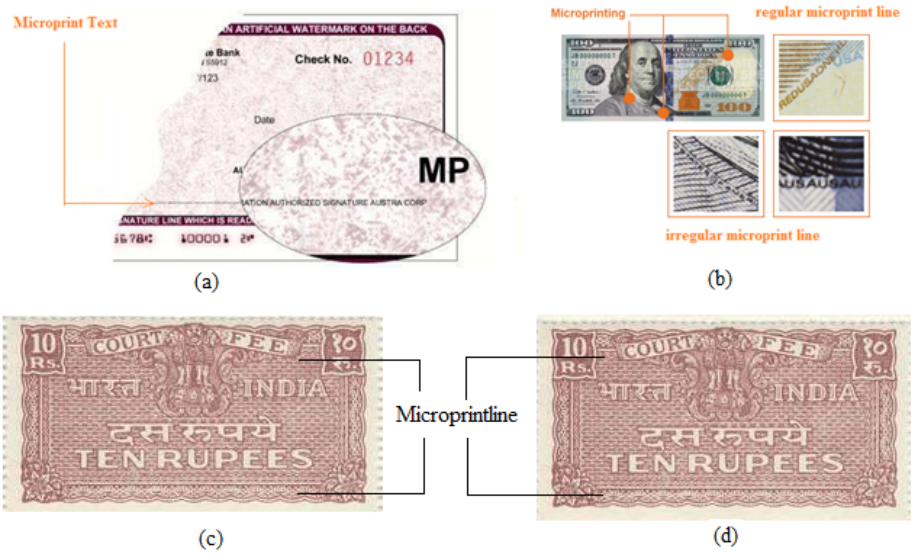



**Figure 2.2:** Micro-print or fine-line printing documents

ferent qualities of the print over paper. There are variations with a particular printing process. For instance, intaglio printing is special kind of offset printing that gives a document a very high quality look that is difficult, if not impossible, to reproduce by using scanners, color copiers or computers with color laser printers. Drying mechanism of intaglio and other-offset printing are quite different. The solvent of offset printing ink is oily based and normally drying under heating effect. Chemical polymerization takes place during drying of offset ink. Here drying time is less and final effect on substrate is brighter and sharper edge. In intaglio printing, drying mechanisms of ink pigment is mainly achieved through evaporation and penetrations of ink solvent on its substrates. Here, the tendency of deposited ink spread beyond its deposition area. In addition, the required drying time of intaglio printing is more than offset printing. Therefore, the final effect of intaglio printing is less bright and less sharp edge as compared to that of offset printing. Examination of final printing effect is an important aspect for verification of security documents' authenticity. Final effects on currency note are defined by unique

color, impression, water resistance, line work (width, thickness, sharpness, etc.), halftone effect, digitized patterns and also reflectivity and feel/tactility. Many of these effects are due to the chemical composition of the ink used and the particular drying mechanism as followed in particular printing technique. Most of these effects are visualized through microscope with high magnification or by different scanning technologies. Forensic experts often take their decision by checking these effects of printing on currency note with mostly the help of a microscope. In this portion of study, we involve machine that closely follows such type approach in order to authenticate the printing process in a bank note.
*Inks-Pigment:* Any printing process involves ink pigment. There are many types of inks like CMYK (cyan, magenta, yellow, and black ink) ink, fluorescence ink, optical ink, etc. Final finish of print result applies varnishing and fused. The basic components of printing ink are pigment, solvent and drier. Where ink pigment is responsible for color effect on a substrate, drier is responsible to bind ink pigment to substrate. Solvent is responsible for soluble of ink pigment and drier. It gives final effect on security documents. particularly, Colour ink pigments contribute to the security documents. Different ink pigments are also used for printing of security documents. For example, fluorescence of color pigment is often used to counter color photocopying of negotiable documents. Fluorescence is normally checked under a UV light source when excited by a specific wavelength. Similarly, key areas of the security design may use fugitive inks that are highly sensitive to a variety of solvents. On the other hand, a relatively new breed of inks known as thermo chromic inks is also used for security measures. Such inks show significant reaction to applied temperature (by changing color or disappearing). They return to their original states upon cooling. Final effects on currency note are defined by unique color, impression, water resistance, line work (width, thickness, sharpness, etc.), halftone effect, digitized patterns and also reflectivity and feel/tactility. Many of these effects are due to the chemical composition of the ink used and the particular drying mechanism as followed in particular printing technique. Most of these effects are visual-

ized through microscope with high magnification or by different scanning technologies. Forensic experts often take their decision by checking these effects of printing on currency note with mostly the help of a microscope. In our study, we involve machine that closely follows a similar approach in order to authenticate the printing process in a bank note.

*Ink-age:* For ink-age determination, several color based features are extracted and change in these features over time has been studied here. The changes of different properties of colors with age are noted. For example, as document is being older much more yellowish and brownish effect on print surface. Dark noise also increases with age. The color shade and brightness are affected on a document with its age and this in turn affects hue and saturation level of the ink colors. Apart from gray level analysis, we consider two other color spaces namely, RGB and HSI. Some color-level features also computed from these color space.

*Security Paper:* The quality of the paper on which the document is printed itself play as security feature. For example, sensitized paper stock reacts differently to use of different solvents and chemicals. Similarly, fluorescent fiber (invisible or visible) based paper is made to fluoresce under excitation from a UV light source. These fibers are photochromic in nature. It spreads randomly on the notes which are illuminated under UV light source. When a banknote is scanned under UV light, the fluorescent paper pulps are visible in the scanned image. The bright spots in the scanned image correspond to the paper pulps present in the note. These pulps play crucial role in authenticating the paper. In a counterfeit note, if the paper is very different from the genuine one, these pulps may not be seen at all. In a high quality counterfeiting, these pulps came as very bright spots and their shapes show significant difference with respect to the pulp marks of the genuine. Therefore, the illumination and shape of these paper pulps are important in characterizing features on a note paper as genuine or fake. In our study we also focused on these features for paper-level authentication.

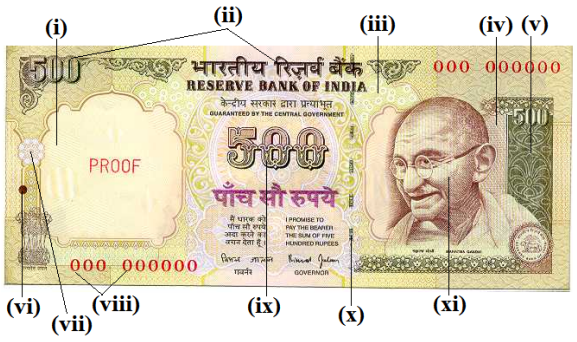
*Security features of Indian Bank Notes:* The impact of the document and its mobility as a transaction medium determines the amount of security in place for its protection. There have been a lot of research studies for providing security in paper documents [80]. Bank notes being the principal monetary exchange medium do have a lot of security features to prevent counterfeit [15]. The Apex bank of every country is responsible for placing these security features in place to prevent forgery. In India it is the responsibility of the Reserve Bank of India (RBI) [78]. The security features in a currency note are mainly on its paper, design and printing. Authentication of currency notes is thus refer to authentication of the following areas: i) Currency notepaper, ii) Currency note printing technique, iii) Ink used for currency note printing, iv) Currency note design, v) Other security features (e.g. the thread, the registration mark, and many others) that are intentionally incorporated to check the authenticity. These security features provide a tough challenge to the counterfeiters who attempt to replicate them. Important security features are associated with the currency notepaper itself. Physical features of currency note are based on its cut size of length, width, grammage and thickness of paper. The paper has a unique feel, crackling sound and it is constituted of high quality with 100% cotton or wood pulp lending a particular color, a unique fiber length, surface finish, a typical opacity, and its capacity of extra strength of folding. Watermarks and security thread are another important parts of security aspect of paper money. Important property of watermark is that it cannot be replicated on scanning or by photocopy equipment. Examination of watermarks checks its design, size and thickness using transparent light. The security thread appears to the left of the Mahatma Gandhi portrait is partially embedded and partially visible. On seeing this thread with an ultraviolet light exposure the thread appears in a single line. This is also a proof of the registration of the note that both the sides of the note are properly aligned. This thread has the writings of "RBI" and "Bharata" in Devanagari) alternatively written on it. Denomination of a note (e.g. 500, or 1000) is also embedded in the thread. Fig. 2.3 shows the significant security

features embedded in both sides of a 500 Rupee note (source: Reserve bank of India). The Guilloche design [79], portrait design, micro lettering, type face, font size, color, see through register, anti-scan lines, Braille mark, rainbow effect, layers of CYMK, bleeding effects, latent image effects, etc. all are involved as part of security design. Beside these, lot of important security features are also involved in the design of currency note which discussed in previous section 1.2.

For proposed automatic authentication, we concentrated above areas for features selection as forensic community concern. Though several features are not included all level of documents. Some are expensive while some may be sophisticated as for requirements. In order to prevent counterfeiting efforts, one way is to embed more expensive and sophisticated security features in documents so that counterfeiting becomes difficult. But inclusion of sophisticated features is not always possible because of the cost of document production. Therefore, for many commonly used security documents, the goal is to embed low cost but easy to authenticate security features.

### 2.3.2 Method of Authentication

For checking authentication of any security document in question, presently so far relied upon the human experts. Forensic departments in different countries have opened up special branch for this purpose. They in general involve different devices (for example, UV lamp, magnifying glass, IR detector, etc.) and observe certain properties of the document in question in order to authenticate the document [78]. However, involvement of such a process is quite difficult if not impossible where a department/office daily has to deal with a huge number of security documents. Moreover, involvement of forensic people requires lots of administrative and legal steps like filing a legal case on finding a document in question, sending the document in original to the police people, waiting for expert's view, etc. making the entire process an unattractive one. Therefore, design of an easy and quick method for authentication of security documents would be of enor-




(a)


(b)

**Figure 2.3:** Significant security features in Indian banknote: **(a)** Front side features: (i) Multidirectional artisan lines, (ii) Intaglio printing, (iii) Omron, (iv) Micro text, (v) Latent image, (vi) Blind mark, (vii) See through, (viii) Fluorescent ink numbers, (ix) Optically variable ink, (x) Security thread with clear text, and (xi) Hand graved portrait; **(b)** Back side features: (i) Multi-directional artisan lines, (ii) Intaglio printing, (iii) Gandhi water mark, (iv) See through, and (v) Omron features.

mous help for the communities that everyday deal with huge number of such documents. Use of a computer based technique could provide a solution to this problem.

A security document, which consists of security features come from several domains. Therefore, exploitation of several security aspects together is needed. This need has been demonstrated by several sub-areas for machine based authentication with different type of security documents. This study is presented a system that computes several low or medium level image features and uses pattern classification technique for detecting different types of features on various sub-domains. In our experimental study, we consider bank cheque first to consider.

Bank cheques are first scanned into color images. Security features are then extracted

from the cheque images. Both kinds of genuine and duplicate cheques are used to generate feature vectors. Distribution of these feature vectors is studied in the feature space and suitable classifier is designed. This is modeled as a 2-class pattern recognition problem, i.e. whether the document belongs to the genuine document class or not. To take this decision, we separately use a support vector machines (SVM) and Neural Network based classification scheme. Let m be the number of genuine samples known as genuine cheques and n be the number of samples known as duplicate. In the feature space, it is expected that these m genuine samples would form a cluster ($C_G$) and n duplicate samples would form another cluster, $C_D$. If $C_G$ and $C_D$ are linearly separable then the task of decision making becomes easier. If D(X) is the linear decision function, then a given cheque, X belongs to $C_G$ if $D(X) > 0$, otherwise X belongs to $C_D$.

For automatic detection of security documents this work proposed a low cost technique for identification. Printing technique identification is one of them. The method attempts to exploit several texture features to identify the printer used to print a document. In this approach, each printed character in a text document is classified using linear or non-linear classifiers. In case of non-linear classifier, it has been trained to distinguish the intaglio printing technique from another one. We conducted an exhaustive study to discriminate the intaglio printing technique from non-intaglio ones and used this research for machine aided authentication of Indian bank notes.

We observed that the research in machine aided authentication of security paper documents is still limited in literature. Paper itself also gives specific identity by its surface finish and crackling sound. During manufacturing process extra features like watermark, security thread and optical fiber (i.e. pulp) are embedded for additional security aspects. The optical fibers or pulps are of specific color and length.. In a counterfeit note, if the paper is very different from the genuine one, these pulps may not be seen at all. In a high quality counterfeiting, these pulps came as very bright spots and their shapes show significant difference with respect to the pulp marks of the genuine. On paper based

authentication we were mainly focused on pulp based authentication. Here, our overall approach is divided into a number of stages: (i) detect pulps in a UV scanned banknote, (ii) extract features from the detected pulps, (iii) train a NN classifier based training samples that include both genuine and fake notes. Once the classifier is trained, we use this for classification which is configured as 2-class (genuine vs. fake) problem. This approach is based on image processing and pattern recognition principles.

Ink age detection sometimes plays crucial role in authenticating a document in question. In our approach we attempt to closely follow the method the forensic experts do in determining ink age. This study attempted automatic ink age detection in a limited domain. pattern recognition techniques for designing automated approach for absolute ink age determination. Basically supervised classifier ANN is used to take a decision. This research assumes several constraints and points out the challenges for future research in this area. For same purpose, we propose a mathematical model on these documents to understand their variation. The main purpose of our model is - 1) to understand the color variation for particular samples OR identify most follower sample on different age segments. 2) Prediction sample's feature after particular time-period 3) testing the proposed model by unknown samples. Nature of this problem is like combinatorial optimization path finding. Ant colony optimization (ACO) [81] based method is best fit for that where ants converge towards a short path, expectantly the optimum or a near-optimum solution towards target. The heuristic information can guide the ants toward the most promising solutions, in a way reminiscent of reinforcement learning. The meta-heuristic ACO explore a much larger number of solutions than greedy heuristics. It efficiently solve the combinatorial optimization problems like highly dynamic network routing applications, NP-hard optimization problems etc. Our proposed model has some similarity with this meta-heuristic ACO.

Finally, one case study has done on real security documents i.e. Indian currency notes. In this study, we have incorporated several other security aspects based on ink, security

thread and an art work based feature set that supplements the printing style based approach. A large dataset comprised has been used to test the system performance and a thorough analysis is provided by comparison with real forensic document examiners to show the applicability of our research in the real world scenario. Analysis is also done to report the speed of the system. Popular pattern recognition tools like the k-means, neural networks and support vector machines are used. We have also shown the robustness of each feature in tackling the problem and thus pointing out the sensitivity of the features. Feature level analysis brings out important issues that the note designing agencies could keep in mind regarding the susceptibility of the features to duplication. A sequential ordering of security features is also suggested that should be maintained while testing to maximize the performance accuracy. Involvement of real samples is another significant aspect of this study. Indian bank notes are used as a reference. Because of highly heterogeneous concepts involved with Indian currencies, authentication of these notes poses a big technical challenge.

#### 2.3.2.1 Classification

Pattern recognition is the method of automatically mapping any input representation for an entity to an output category. The recognition task is generally categorized based on how the learning procedure determines the output category. This learning procedure can be supervised, unsupervised and semi-supervised. In supervised learning procedure a given pattern is assigned to one of the pre-defined classes, using labelled data to build a model or guide the pattern classification, Whereas unsupervised learning procedure a given pattern is assigned to an unknown class and In semi-supervised learning procedure a given pattern is assigned to one of the pre-defined classes, using both labelled and unlabeled data. For supervised learning, or classification, a functional model is often used to map observed inputs to output categories. A great deal of model construction techniques have been developed for this purpose [82], including decision trees,

rule induction, Bayesian networks, memory-based reasoning, Support Vector Machines (SVMs), and neural networks.

The experimental samples of our work are tested by different classification procedures. All the samples are at first clustered using some unsupervised clustering method. The purpose of this clustering is to analyze the distribution of samples in the feature space. Next, non-linear classifiers are evaluated by computing the number similar samples grouped together vs. the number of dissimilar samples contained in that group. Since all samples are tagged with their classes evaluating clustering results in this way is straightforward.

*Classification using k-means:* The K-means algorithm is a popular data clustering algorithm. To use it requires the number of clusters in the data to be pre-specified. Finding the appropriate number of clusters for a given data set is generally a trial-and-error process made more difficult by the subjective nature of deciding what constitutes correct-clustering.

For this purpose, we implement a k-means algorithm. The k-means clustering is done to analyze the linear distribution of samples in the feature space. Suppose, k-means clustering aims to partition N observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. Given a set of observations $(x_1, x_2, ..., x_N)$ where each observation is a d-dimensional real vector, k- means clustering aims to partition the n observations into $k(<= N)$ sets $S = S_1, S_2, ..., S_N$ so as to minimize the within-cluster sum of squares (WCSS). In other words, its objective is to find :

$$argmin_S \sum_{i=1}^{k} \sum_{x \in S_i} \parallel x - \mu_i \parallel^2 \tag{2.1}$$

Where,$\mu_i$ is the mean of points in $S_i$. Since all samples are tagged with their classes evaluating clustering results in this way is straightforward.

If the algorithm finds two clusters one corresponding to genuine samples and another for duplicate samples then the value of k set to 2. Initialization is done by choosing two samples randomly to initialize two cluster centres. The k-means results are evaluated by computing the number of similar samples grouped together vs. The number of dissimilar samples contained in that group. Where, all samples are tagged with their classes i.e. genuine or duplicate for evaluating clustering.

*Classification using SVM:* Machine learning and pattern recognition communities have been frequently used Support Vector Machines (SVMs) [82], [83] which are used for constructing supervised classification models for high-dimensional representations. The observations from the training set which best define the decision boundary between classes are chosen as support vectors. Particularly, each support vector is assigned by non-zero weights. The classification of a test object is predicted by computing a weighted sum of similarity function output between the new object and the support vectors for each class. The predicted class is the class with the largest sum, relative to a bias (offset) for the decision boundary. Different Kernel functions are used to measure the similarity between objects. Cosine similarity function and the Gaussian RBF are common kernel functions for this purpose. SVMs have been used for time series classification by several researchers. Researchers have successfully been applied to several different areas ranging from face-recognition and verification, speaker verification, text categorization, prediction, image retrieval, and handwriting recognition.

So,support vector machines (SVM) are used aiming at determining the location of decision boundaries that produce the optimal separation of classes. Two different types of nonlinear kernel functions namely polynomial and radial basis function (RBF) were used in our experimental study. These two kernel functions are defined as:

$$polynomial : k(x, x') = k(x.x' + 1)d \tag{2.2}$$

$$RBF : k(x, x') = exp - \gamma(x - x')^2, \gamma > 0 \tag{2.3}$$

Where x denotes training vectors and d, $\gamma$are kernel parameters. Mean squared error (MSE) is computed as follows:

$$\text{MSE} = \frac{\sum_{j=1}^{T}\sum_{i=1}^{V}(d_{ij} - y_{ij})^2}{V \times T} \tag{2.4}$$

where $V$ and $T$ are the number of support vectors and test samples, respectively, $d_{ij}$ and $y_{ij}$ are the desired output and the SVM output, respectively for the $i$-th support vector and the $j$-th test sample.

*Classification using Neural Network:* Neural networks have been intensively used in the area of pattern recognition and have increasingly received attention in various areas such as signal processing, pattern recognition and automatic control. Artificial Neural Networks (ANN) [84], also known as neural networks or neural nets, are analogous to their namesake, biological neural networks, in that both receive multiple inputs and respond with a single output. ANNs are comprised of connected nodes, (i.e. neurons), which are partitioned into layers. Networks may enclose an input layer, an output layer, and a hidden inner layer, and additional hidden layers may be added to increase the complexity of the network. Weights are assigned to each of the links between neurons, and they are updated as part of the learning process. A simple example of an ANN is a perceptron [85], which is a two-class linear classifier that is composed of a single-layered neural network. During training, each example is fed through the network, and the weights are updated according to the difference between the actual and expected output. During testing, the weighted sum of the network (the dot product of the instance and the vector of weights) is compared to a bias parameter to classify each instance as either a positive or negative instance. ANNs generalize well to previously unseen instances because they map every possible input to some output. The most popular network is the multilayer perceptron architecture which is trained using the back-propagation algorithm. It consists of at least three layers: an input layer, an output layer and one or more hidden layers. Additionally, they are robust because even if a node is faulty or is removed, since

the work is distributed across the network, it is still possible to achieve good results. ANNs have been successfully applied to domains such as character recognition [86], face detection [87], intrusion detection [88], speech recognition [89], autonomous driving [90], and astronomy [91]. The classification accuracy of the tested samples is also checked with a Neural Network (NN)-based classifier. Particularly, MLP (Multi-Layer Perceptrons) based NN structure used in our work. The generalized function of NN is:

$$y = \sum_{j} \omega_j \phi_j(x) \tag{2.5}$$

Where $\omega_j$ are the weights (which are updated following the K-means optimization techniques) and $\omega_j(x)$ is the transfer or activation function. For documents authentication multi layer perceptron nural network (MLP) is used to design a Neural Network-based classifier. Well-known back propagation algorithm is used to train this network. The network does use of the following logistic function as transfer or activation function.

$$f(x) = \frac{e^x}{1 + e^x} \tag{2.6}$$

Another activation function i.e. Gaussian Radial Basis Function (RBF) also used as follow:

$$\phi_j(x) = exp(- \parallel x - c_j \parallel^2 / 2\sigma^2) \tag{2.7}$$

$c_j$ represents the center of j-th cluster, $\sigma$ is the $\parallel x - c_j \parallel^2$ is the square of the distance between the input feature vector x and the cluster center for the radial basis function. A popular gradient descent method is used to find the optimized set of connection weights that are updated as per the following equation. Update of weights does consider both the t-th and $(t-1)_{st}$ weights to compute the $(t+1)_{st}$ weight as follows.

$$W_{t+1} = W_t + \alpha * (\frac{\partial E}{\partial W})|_{W(t)} + \beta * (W_t - W_{t-1}) \tag{2.8}$$

Where $\alpha$ is the learning parameter, $\beta$ is known as the momentum and E is the error term which is calculated as $E{=}\frac{1}{2}(T-O)^2$ ; Where $T$ is Target and $O$ is Output.

#### 2.3.2.2 Feature reduction

Linear Discriminant Analysis (LDA) [92–94] is a data driven technique looking for linear transformation allowing for dimensionality reduction of features. The aim of LDA is to preserve information important for discrimination between feature vectors belonging to different classes. Therefore, for each training feature vector, we need also information about the class to which the vector belongs. LDA allows to derive linear transformation with bases sorted by their importance for discrimination between classes. Therefore, for the purpose of dimensionality reduction, we can project features only into several first basis, which preserve almost all the variability in data important for the discrimination of classes. LDA also ensures the de-correlation of features. Moreover, it does not de-correlate only overall training data, but features belonging to each particular class are also de-correlated. However, assumption that features belonging to each particular class obey Gaussian distribution and that all the classes share the same covariance matrix must be fulfilled for the optimal functionality of LDA. However, classes are well separated in the direction corresponding to the first LDA base vector. Linear discriminant analysis (LDA) and its related Fisher's linear discriminant are methods used in statistics and machine learning to find the linear combination of features which best separate two or more classes of objects or events. The resulting combination may be used as a linear classifier, or, more commonly, for dimensionality reduction before later classification. Fisher linear discrimination analysis (LDA) for authentication can derive by following steps.

Suppose $\vec{\omega}$ is the normal to the discriminating hyper plane, $\vec{\mu}_{y=0}$, $\vec{\mu}_{y=1}$, are the sample means, $\sum_{y=0}$ and $\sum_{y=1}$ are the covariance matrices of the two classes (*genuine* and *fake*). Now we can consider that for one class, $y = 0$ for $\vec{\omega}^T.X + b \geq 0$ and for the other

class, $y = 1$ for $\vec{\omega}^T.X + b < 0$. From these equations, parameter vector $\vec{\omega}$ is computed to maximize class separability criterion and b is the bias, which lies in between the means of the training samples projected onto this direction. The separation between these two distributions is to be the ratio of the variance between the two classes and is given by

$$S = \frac{\sigma^2_{\text{between}}}{\sigma^2_{\text{within}}} = \frac{(\vec{\omega}.\vec{\mu}_{y=1} - \vec{\omega}.\vec{\mu}_{y=0})^2}{(\vec{\omega}^T \sum_{y=1} \vec{\omega} + \vec{\omega}^T \sum_{y=0} \vec{\omega})} \tag{2.9}$$

For Fisher LDA, this separation achieves maximum when

$$\vec{w} = (\sum_{y=1} + \sum_{y=0})^{-1}(\vec{\mu}_{y=1} - \vec{\mu}_{y=0}) \tag{2.10}$$

Finally, the cumulative accuracy taking all the features into consideration is computed and a study is conducted to find out a proper sequence in which the features are to be tested. This is important as this reduces the load of the machine if at various stages the number of notes to be checked could be reduced without sacrificing the accuracy.

## 2.4 Summary

Criminal activity concerning generation of duplicate copies of certain kinds of security documents has become a potential threat to our civil society. Our works provides an automatic means for verification of authenticity on security documents. In this chapter we provided over view of different framework with their method of solution. Our entire proposed framework considers design of an efficient but low-cost solution to this problem of detecting duplicate documents so that mass scale deployment of such systems can be feasible. Next following chapter we discuss every pointed method and experimental result elaborately.

# Chapter 3

# Security Design Verification

*Note: The work of this chapter is related under publications [4], [5] and [6].*

## 3.1 Introduction

Many features are available for embedding security in important paper documents [7] [95]. These features vary in cost of embedding as well as their potential in retaining authenticity of documents. Among these features, light fine-line based security patterns are commonly appear on background of security documents. Two popular techniques are used to build these patterns one is micro-print line and another is line half-tone artwork. This type of artwork is difficult to reproduce on scanning equipment or replicate by other printing methods. Simultaneously, authentication proof is also importance for present era.

*Microprint line :* Microprint-line based design is used in large number of security paper documents because of its low cost and high capability in providing security to paper documents. The frequent use of microprint line design is based on the fact that this is clearly readable in authentic documents under magnification, but it becomes blurred and unreadable when copied or scanned [96], [97]. Therefore, such a security feature is used in majority of security documents including bank cheques, legal deeds, certificates,

mark sheets, postal stamps, etc. The use of an intricate microprint-line background design is commonly introducing by Guilloche module [79]. Latten image and portrait type of images is also constructed by this micro-print artwork. Apart from using fine-line artwork, micro-sized typeset characters can also be included in the design. These types of patterns are generally embedded in currency notes, bank cheques, legal deeds, certificates, mark sheets, postal stamps, etc. These security documents frauds are on rise as sophisticated reproduction machineries are available in market.

So, authentication checking of microprint line security features also is in question. Presently, it is relied upon the human experts. However, involvement of their process is quite difficult if not possible where a department/office daily has to deal with a huge number of such documents. Moreover, involvement of forensic people requires lots of administrative and legal steps. Therefore, design of an easy and quick method for authentication of microprint-line security features would be of enormous help for the communities that everyday deal with huge number of such type of security documents. This study addresses the above problem and proposes an automatic approach for detecting duplicates while processing security documents. A financial bank is a nice example that daily deal with huge number of security documents in form of processing of cheques ( figure 3.1), drafts, pay-orders, etc. The approach is based on pattern recognition principles by which relevant features are initially extracted from cheques and then using these features an algorithm for discrimination between genuine and duplicate cheques is outlined.

*Halftone :* Halftone images are essential part of printed materials [98] [99] [100]. In printing pictures in books, a continuous tone original picture is first converted into a halftone (HT) one which finally gets printed in the book. There are two types dots by which halftone images are composed of: (i) dispersed dot and (ii) clustered dot. Dispersed dots are of fixed size and dot diameters are not directly related to the dot frequency. The number of dots in a region defines the basis of tonal levels. The clustered dot occurs

when the halftone dots are of variable sizes. In this scheme, the dot diameter is proportional to the dot frequency. Dispersed dots are used in limited case digital printing (e.g. laser jet printers, photocopiers, etc.) whereas clustered dots are used in large scale printing (e.g. offset printing, lithography, silkscreen printing, etc.).

The clustered dots can be of different shapes. The line halftone dots are one of them and mostly used by the printing houses. Line half tone images are commonly used in printed books, old manuscripts, magazines etc. including many security documents like certificates, bank checks, currency notes, legal deeds and so on. In security documents, line halftone images are normally used as background design that serves as a protection against counterfeiting. Such design involves micro print-line patterns, guilloches patterns, latent image pattern, relief line pattern etc. Most of these patterns are produced from continuous tone image. The design details of such patterns are not clear with the naked eye but become clear with magnification. Characteristics of these line patterns are line thickness, line density and ink colour. Fine-line design features are changed in the event of a photocopying attack. For example, when a forger attempts to copy the page, the design will appear blurred and display a pattern spread. Generally, a document examiner inspects this deformation with a magnifier [62], [7], [101]. The document in question is inspected using different light sources, i.e. transmitted light, oblique light, etc. This inspection is grossly manual and therefore, time consuming. For quick decision-making and for better visual inspection, a sophisticated machine assisted technique is called for. This study is aimed at developing an inverse half toning technique (IHT) for authenticating line HT images.

The method attempts to formulate a statistical measure in order to judge the quality of the image in question against the original image. We have considered line halftone image at different resolutions namely, line per inch (lpi) which are commonly used in practice. The significant contribution of this study is to use learning based pattern classification technique for designing the IHT method. The existing methods rarely exploit this tech-

nique rather make use of static template or edge analysis based pattern matching. Many techniques borrow idea from digital signal processing. Pattern classification based inverse halftoning has also been attempted in few works [102], [103] . These methods do not consider resolution of an input image separately and therefore, inverse halftoning is done based on a overall learning over images of many resolutions. Our method brings novelty by finding the lpi information so that generation of inverse halftone becomes more precise. Secondly, empirically we observe that use of more than one neural net gives better quality inverse halftone than the one given by only one neural net.

## 3.2 Artwork based Authentication Proposal

On both experiments are making out by pattern recognition principal. For bank cheques recognition, both linear and non-linear classifiers are considered for designing classification module. But in 2nd experiment is aimed at developing an inverse half toning technique (IHT) though neural network (ANN) for authenticating line-HT images. The method attempts to formulate a statistical measure in order to judge the quality of the image in question against the original image. Next two subsections we discussed these proposed microprint/fine-line base authentication techniques and IHT elaborately.

### 3.2.1 Microprint-line

#### 3.2.1.1 Features Extraction

On subsection 2.3.1 we have already discussed on the domain of art-work features for document authentication. As mentioned earlier that bank cheques are taken as a reference in this study to validate the proposed method. Professionals from printing press are requested to design sample cheques for our study. All the security measures that are normally taken care of during designing and printing of real bank cheques are also followed here in designing sample bank cheque. Prints of cheques are taken from a printer

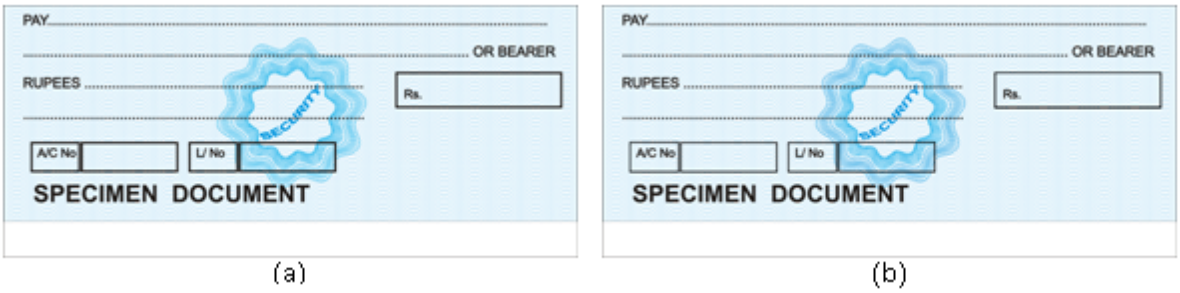



**Figure 3.1:** Image of bank cheque samples: (a) a genuine cheque and (b) a duplicate cheque

specially designed for security printing. Fake or duplicate cheques are generated using the means that are commonly used in forgery of such documents. Questioned document examiners in the field of forensic sciences are consulted for this purpose to know the process of forgery in details. Hi-end scanners, printers (mostly offset), sophisticated image editing software packages [62] and experts, etc. are involved in generating duplicate versions of bank cheques. The kind of paper used for taking print out is very similar in quality to ones used for printing of cheques in press. Finally, the duplicate cheques after printing are manually checked to verify their perceptual similarity with the authentic ones. Computation of the features requires registration of two images $d_R$ (for reference image) and $d_T$ (for tested image). The registration has been done based on the four control (or tie) points. These control points are chosen from the set of pixels whose locations in the two images are known. In this section of study, the right side rectangular box ( figures 3.1), i.e. the box for writing the amount in numerals is considered for registration. Detection of this box in the images is rarely missed and this process needs less amount of computational effort. Once this box is detected, its four (outer) corner points are chosen as control or tie points for registration.

The security features in documents like bank cheques are also embedded on three areas namely (i) color features, (ii) background artwork and logo, and (iii) paper quality. In this section of study, we capture features related to the first two categories. Four different aspects are investigated for color related security and two more features are computed

from the background artwork or texture patterns. Computation of these six features and rationale behind choosing them to discriminate genuine vs. duplicate documents are discussed below.

*Average Image Hue ($f_h$):* In printing color theory, hue defines the quality of a color. Perceptually two colors may look alike but if they are printed with different ink pigmnets they will occupy different position in the color cube. Actually, dominant wavelength is a physical analog to this perceptual attribute, i.e. hue. Because of this reason, comparison of hue values in two images may give a significant clue to decide whether color quality in those images are same. Figure 3.2 (a) and (b) demonstrate the histograms of hue values for the genuine and duplicate cheuqes corresponding to the images in figure 3.1 (a) and (b). Since the proposed system takes an RGB image (of documents, e.g. bank cheques) as input, hue for an individual pixel $(p), h_p$ is computed from its RGB values $r_p, g_p$ and $b_p$ as follows:

$$h_p = \begin{cases} \theta & if b_p \leq g_p \\ 360 - \theta & if b_p > g_p \end{cases} \tag{3.1}$$

where $\theta$ is the angle measured with respect to the red axis of the HSI color space. For each pixel, $h_p$ is measured and then an average hue is computed for the entire image. Let $f_h$ denote this average hue.

*Gray level variation ($f_{gv}$):* The standard deviation of the gray level distribution (of the image pixels) is considered as a feature and denoted by $f_{gv}$ :

$$f_{gv} = \sqrt{\frac{\sum (g_p - \bar{g})^2}{N - 1}} \tag{3.2}$$

where N is total number of pixels, $g_p$ be value of the gray value of pixel p and $\bar{g}$ be the mean gray value of the image. Figure 3.2 (c) and (d) demonstrate the histograms of gray values for the genuine and duplicate cheuqes corresponding to the images in figure 3.1 (a) and (b). Please note that a color image is transformed into its corresponding

gray image by using the perceptually best-known weighting factors (i.e. Gray = 0.3 Red + 0.59 Green + 0.11 Blue).

*Binary correlation ($f_{bc}$):* Computation of this feature assumes the knowledge of authenticity (or genuineness). A given (or target) document ($d_T$) is compared with the genuine one ($d_R$). Here a correlation between two binary images (gray images are converted into binary images using Otsu's thresholding method [104]) is measured. The correlation coefficient (i.e. a similarity measure), r between the reference (i.e. genuine) and the target image is measured as,

$$r(d_R, d_T) = \frac{1}{2} - \frac{s_{10}s_{01} - s_{00}s_{11}}{2\sqrt{(s_{11}+s_{10})(s_{01}+s_{00})(s_{11}+s_{01})(s_{10}+s_{00})}} \tag{3.3}$$

where $s_{00}, s_{11}, s_{01} and s_{10}$ denote the number of zero matches, one matches, zero mismatches, and one mismatches, respectively. This coefficient lies in [0, 1] and gives the value for the feature, $f_{bc}$.

*Kurtosis of image colors ( $f_{kr}, f_{kg} and f_{kb}$ ):* Apart from measuring standard deviation of gray values we also measure kurtosis to analysis whether variations is due to infrequent extreme deviations. R, G, and B channels are separately considered to measure respected kurtosis. For instance, let $f_{kr}$ denote the kurtosis of red channel and it is measured as

$$f_{kr} = \left\{ \frac{N(N+1)}{(N-1)(N-2)(N-3)} \sum \left( \frac{r_p - \bar{r}}{\sigma_r} \right)^4 - \frac{3(N-1)^2}{(N-2)(N-3)} \right. \tag{3.4}$$

where $r_p$ be red value of pixel p, $\bar{r}$ and $\sigma_r$ be the mean and standard deviation of red values of the image. Similarly, $f_{kg}$ and $f_{kb}$ are computed to denote kurtosis of green and blue channels, respectively.

The following two features are considered to investigate the changes in the background artwork due to forgery. This artwork is basically a texture pattern and if the forgery were done very carefully it would be difficult to identify the changes with open eyes.

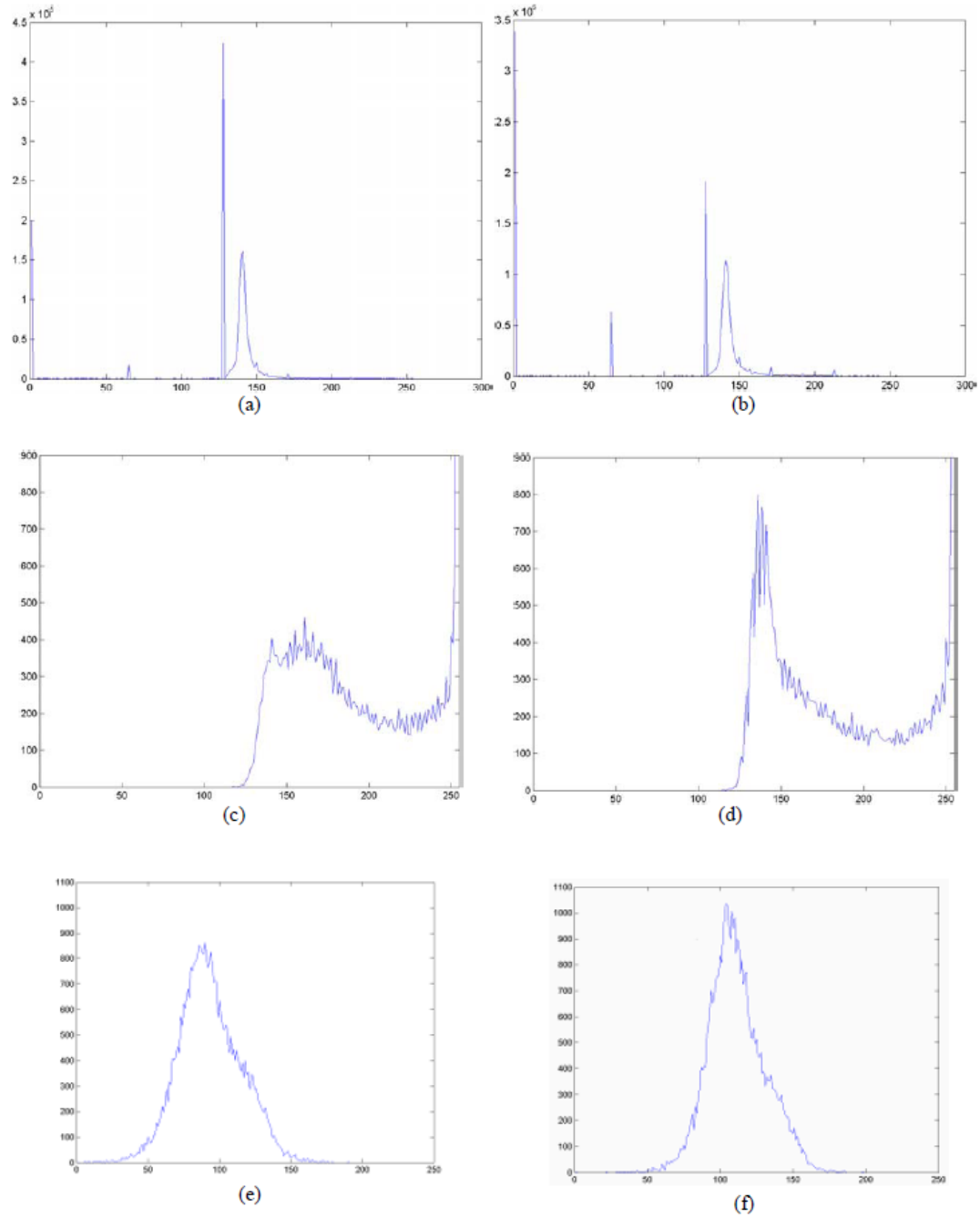



**Figure 3.2:** Artwork features: (a) and (b) hue values (c) and (d) gray values (e) and (f) centeredFourier spectrum by log transformation corresponding to the genuine and duplicate cheques infigure 2(a) and (b), respectively

Forensic people use magnifying glass or microscope to detect changes, if any. In the present study, features are also carefully extracted so that they would be able to report the unexpected changes, if any exists, in a given document image.

*Measure of Line Quality ($f_l$):* Since the background artwork is basically a line drawing, it has been experienced by the forensic community that scanning and subsequent printing of the scanned image results in broken lines in the artwork part (though perceptually the artwork may look like one in the genuine document). To capture this attribute, we measure the line quality of a given image as follows.

In the binary image, straight lines originating from a pixel in three directions namely horizontal (left to right), vertical (top to bottom) and diagonal (in the South-East direction) are identified. Length of each such line is recorded. Isolated pixels are not counted. Detection of such lines started from the leftmost top pixel and the algorithm scans pixels in row major order. Once a line is identified, its containing pixels are (virtually) deleted so that the same straight line or its part is not counted repeatedly.Let L be the number of lines identified in the above process and $l_i$ be the pixel length of the i-th line. The average length of these lines gives the value of the feature, $f_l$ and is computed as,

$$f_l = \frac{1}{L} \sum_{i=1toL} l_i \tag{3.5}$$

*Fourier Power Spectrum* ($f_{ps}$) : Since the background artwork is a result of repetitive pattern (or texture), Fourier analysis of the image provides significant clues to identify unexpected changes in the artwork pattern. Gray version of the image is considered and its centered Fourier Spectrum is investigated. In this spectrum as the dominant frequencies are distributed over a very small region around the center, it's difficult to differentiate a nicely forged duplicate document from its corresponding genuine copy. However, if we consider the log transformation of this spectrum, differences become more evident and significant. Figure 3.2 (e) and (f) demonstrate this fact. In fact, the Fourier spectrum for the genuine and duplicate bank cheques corresponding to figure

3.1 (a) and (b) look similar (not shown here), however, when their log transformations are considered differences between them become more clear as shown in Figure 3.2 (e) and (f). This attribute is captured in the feature $f_{ps}$ and computed as-

$$f_{ps} = \log(1 + FS) \tag{3.6}$$

where $FS$ is the centered Fourier spectrum, $fs = \sqrt{R^2 + I^2}$. R and I are the real and imaginary parts of the Fourier Transform, F.

#### 3.2.1.2 Microprint-line base authentication method

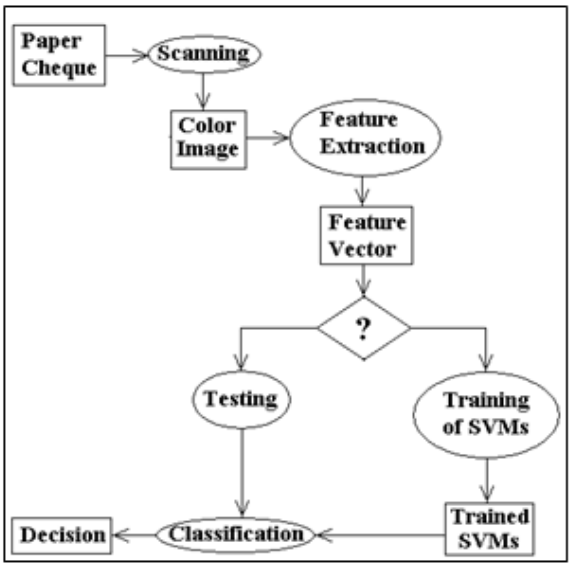



**Figure 3.3:** Schematic diagram of the proposeddocument authentication system

Here, figure 3.3 shows a schematic diagram of the proposed approach. Bank cheques are first scanned into color images 3.1. Security features are then extracted from the cheque images. Both kinds of genuine and duplicate cheques are used to generate feature vectors. Distribution of these feature vectors is studied in the feature space and suitable classifier is designed.. After the features which described on previous section 3.2 are

extracted from a bank cheque document image. For its authentication, this recognition model based on 2-class pattern recognition problem, i.e. whether the document belongs to the genuine document class or not. So, overall recognitions are followed as on earlier section 2.3.2.1. For the present study, support vector machines (SVM) and neural network (ANN) are considered for designing classification module. A set of labeled samples is used to train the SVMs or ANN. Next another set of samples is used to verify the classification accuracy. Experiment is conducted to confirm that the security features captured computationally are sufficiently robust to discriminate authentic vs. duplicate document.

For this defined purpose, we implement first k-means algorithm and cluster the two different labelled samples into two classes. Selecting two samples randomly initializes the centers in k-means algorithm. Since k-means results get affected by this initialization phase, k-means is executed more than once (three times) and each time the clustering results are investigated. This investigation reveals that the clusters always overlap and therefore, it is difficult to find D(X), i.e. a linear decision function.

The classification accuracy is then checked with a Neural Network (NN)-based classifier. An MLP (Multi-Layer Perceptrons) based NN structure used. The MLP consists of 8 input nodes correspond to eight dimensions of a feature vector. The output consists of only one node to gives binary output ('1' indicates authentic and '0' indicates duplicate). Hidden layer, in the present experiment, contains 2 nodes. A logistic function as explained in previous section 2.3.2.1 is used as the activation function of this network. Update of weights done by gradient descent method as discussed in the previous section 2.3.2.1. Like SVM-based classifier a four-fold test is conducted for NN-based classification. The proportion in which samples appear in training, validation and test data is as same as they were in case of SVM, i.e. 2:1:1 (i.e. training: 50%, validation: 25% and testing: 25%).

#### 3.2.1.3 Microprint-line base authentication result

As it mentioned earlier that bank cheques are taken as a reference in this study to validate the proposed method. Professionals from printing press are requested to design sample cheques for our study. All the security measures that are normally taken care of during designing and printing of real bank cheques are also followed here in designing sample bank cheque. Prints of cheques are taken from a printer specially designed for security printing.

Fake or duplicate cheques are generated using the means that are commonly used in forgery of such documents. Questioned document examiners in the field of forensic sciences are consulted for this purpose to know the process of forgery in details. Hi-end scanners, printers (mostly offset), sophisticated image editing software packages [5] and experts, etc. are involved in generating duplicate versions of bank cheques. The kind of paper used for taking print out is very similar in quality to ones used for printing of cheques in press. Finally, the duplicate cheques after printing are manually checked to verify their perceptual similarity with the authentic ones.

In total, two hundred bank cheques (one hundred genuine, one hundred duplicate samples) are considered in the present study. A flatbed scanner is used to scan these cheques at 300 dpi, true color images, which are stored as uncompressed .tiff image format. Each cheque takes about 16M storage spaces.

*Result of k-means* All the samples are at first clustered using some unsupervised clustering method. The purpose of this clustering is to analyze the distribution of samples in the feature space. The k- means algorithm is used for this purpose which described in previous chapter section 2.3.2.1 The algorithm finds two clusters one corresponding to genuine samples and another for duplicate samples, i.e. value of k is set to 2. Initialization is done by choosing two samples randomly to initialize two cluster centers.

The k-means results are evaluated by computing the number similar samples grouped together vs. the number of dissimilar samples contained in that group. Since all samples

are tagged with their classes (genuine or duplicate) evaluating clustering results in this way is straightforward. Table 3.1 presents the evaluation of k-means results. Since cluster centers are initialized randomly, k-means were executed three times to get an average result.

From Table 3.1, overlapping of samples in the feature space can easily be visualized. Therefore, it would be difficult to find any linear decision boundary to separate the genuine and duplicate samples. This results in discarding design of a classification system based on any linear decision functions.

**Table 3.1:** K-means results for clustering of samples in two clusters

| | Distribution of samples in clusters | | | | Clustering Accuracy |
|---|---|---|---|---|---|
| | Samples in Genuine (G) | | Samples in Genuine (F) | | |
| | G | D | D | G | |
| Iteration 1 | 90 | 12 | 88 | 10 | 174(87 %) |
| Iteration 2 | 88 | 14 | 86 | 12 | 174(87 %) |
| Iteration 3 | 90 | 10 | 90 | 10 | 180 (90 %) |
| Average | 89.3 | 12 | 88 | 10.7 | 177.3 (88.7 %) |

*Classification using SVM* Support vector machines which described in previous chapter under section 2.3.2.1 are incorporated here to design a classification scheme. Two different types of non-linear kernel functions namely polynomial and radial basis function (RBF) were used in the present experiment. The set of 200 samples are divided into four sets to realize a four-fold experiment. In each run of an experiment, two sets are considered as training sets, the remaining two sets serve as validation and test sets. Sets are selected in such a way so that each set appears at least once as a test set and a validation set. To ensure that each set would eventually appears twice as training set and four different runs were executed. The result of this four-fold experiment is reported in Table 3.2. It is to be noted that the following values were estimated for the kernel parameters for polynomial: d = 3 and for RBF: $\gamma = 1$ and they do not change with changing of training sets. Table 3.2 shows some important observations. For the present problem, polynomial kernel function performs better than an RBF based kernel.

Moreover, number of support vectors used by a polynomial kernel is far less than the number used by a RBF based kernel. However, values for mean squared error (MSE) show that RBF kernel gives very low MSE when compared to a polynomial kernel.

**Table 3.2:** Classification of bank cheques using SVM

| | Support Vectors | | Classification accuracy on Test Dataset | | MSE | |
|---|---|---|---|---|---|---|
| | Polynomial | RBF | Polynomial | RBF | Polynomial | RBF |
| Run1 | 6 | 65 | 100 | 100 | 4.806 | 0.06 |
| Run2 | 7 | 56 | 96 | 96 | 0.982 | 0.169 |
| Run3 | 5 | 61 | 98 | 96 | 2.05 | 0.154 |
| Run4 | 5 | 60 | 98 | 96 | 1.139 | 0.169 |
| Avg. | 5.75 | 60.5 | 98 | 97 | 2.24 | 0.138 |

*Classification using Nural Network:* As mentioned in previous chapter section 2.3.2.1 an MLP is used to design a Neural Network-based classifier. Well-known back propagation algorithm is used to train the network. In the present experiment, $\alpha$ is set to 0.9 and $\beta$ is assigned 0.1. The same dataset as used for SVM-based classifier is also used here to train and test the MLP. Like SVM a four-fold experiment is conducted and results are reported in Table 3.3. The training, validation and test sets used in different runs of experiments are exactly the same for designing SVM as well as NN-based classification scheme.Table 3.3 shows NN-based classification gives about 97.5 % accuracy in classifying test documents as genuine or duplicate. The accuracy is slightly less than that of SVM-based classifier but both the results are definitely comparable.

**Table 3.3:** Classification of bank cheques using Neural Network

| | Classification of bank cheques using Neural Network | | | |
|---|---|---|---|---|
| | Training Dataset | | Test Dataset | |
| | Sample | Correct Classification | Sample | Correct Classification |
| Run1 | 100%(G: 54, D: 46) | 100% (G: 54, D: 46 | 50(G 23, D 27) | 98 (G 23, D 26) |
| Run2 | 100%(G: 46, D: 54) | 98% (G: 45, D: 53) | 50(G 30, D 20) | 98 (G: 29, D: 19) |
| Run3 | 100%(G: 60, D: 40) | 98% (G: 60, D: 38) | 50(G 22, D 28) | 98 (G: 22, D: 27) |
| Run4 | 100%(G: 52, D: 48) | 97% (G: 51, D: 46) | 50(G 20, D 30) | 98 (G: 19, D: 30) |
| Avg. | | 98.25 % | | 97.5 % |

### 3.2.2 Halftone image authentication

#### 3.2.2.1 Inverse Halftone method

The main of the inverse transform method is to take a line halftone (HT) image as input and produce a high quality gray-scale image corresponding to the input. A radial basis function neural net (RBF-NN) is the core of this transform method. The reason for using RBF-NN as the neural network lies in the fact that the inverse transform is a complex non-linear process and for doing this RBF-NN shows better performance for universal non-linear approximation over the other neural nets (e.g. MLP-NN) [102]. The reconstruction function is given by-

$$C = \sum_{h}^{n} w_h \cdot \phi(||x - t_h||) + B \tag{3.7}$$

where $n$ is the total number of input samples applied for output neuron $C$ which corresponds to the intensity level of a pixel $(i,\ j)$ in the output image, $w_h$ is the synaptic weight connecting hidden neuron $h$ to output neuron, $B$ is a bias of the output neuron and the activation function $\phi(\cdot)$ is defined as-

$$\phi(||x - t_h||) = \exp\left\{\frac{-(||x - t_h||^2}{2\sigma_h^2}\right\}, \tag{3.8}$$

where the set of centres $\{t_h|h = 1, 2, \ldots, n\}$ are $m^2$-dimensional vectors to be determined, $x$ is the $m^2$ dimensional pattern obtained by placing a $m \times m$ template around the $(i,\ j)$ pixel of the input image, and $\sigma_h$ is the variance of Gaussian function. The gradient descent is used for error-correction learning process.

*Architecture 1.* A single RBF-NN is used that takes the binary HT image as input, predicts lpi of the input image and generates the gray image as output. This is somewhat similar to what has been used in previous works for inverse halftoning of dithered halftone images [102] but these works do not compute lpi information.

*Architecture 2.* This architecture consists of two levels of RBF-NNs as shown in Figure 7.2. The first level takes the input image, $I_{HT}$ and predicts the lpi information for the given image. In addition, the first level produces a low level ($k$-level) image, $I_k$. The lpi information helps to choose a particular NN on the second level. If we consider three different lpi based HT images ($I_{HT}$) then three different RBF-NNs corresponding to three specific lpi values are present on the second level. Depending upon the lpi detected by the first level RBF-NN, the intermediate $k$-level image ($I_k$) is passed to the particular RBF-NN on the second level. The RBF-NN of the second level produces the final gray scale image ($I_G$). The values of $m$ and $k$ have definite impact on the quality of the inverse halftone and therefore several alternatives have been tried. In our experiment, three different values of $m$ namely, 3, 5 and 7 and three different values for $k$ namely, 4, 8 and 16 are used. Variation in m will give different pixel templates based on which prediction of lpi information and the intensity level of a pixel in the output image is predicted. Variation in $k$ defines the intermediate approximated image at different intensity levels.

*Dataset:* In this experiment five standard digital grayscale images namely, (i) *Peppers*,

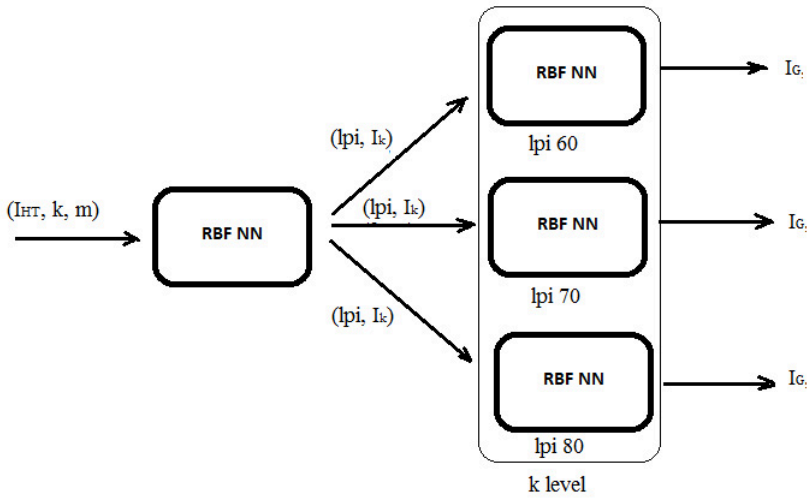



**Figure 3.4:** Block diagram of Architecture 2

(ii) *Mandrill*, (iii) *Barbara*, (iv) *Atlas hand* and (v) *Lena eye* which have no background have been considered. Line HT images are generated using a commercial software namely Adobe Photoshop Software 7.0. All the images are processed 100 dpi resolutions with specified screen angle namely, 45° and dot frequencies of 60, 70 and 80 lpi. Printing is done through single color offset printing machine (black ink used here). Printed images are digitized by flatbed HP scanner (ScanJet 8250) with same resolution (i.e. 100 dpi). Binary images are obtained by using Otsu thresholding method. HT binary images of the first three images (i.e. *Peppers*, *Mandrill* and *Barbara*) have been considered for training the RBF-NNs and remaining two images (i.e. *Lena eye* and *Atlas hand*) have been considered for testing. For architecture-1, i.e. use of only one RBF-NN, about 500,000 (500K) binary feature vectors tagged with lpi information and gray value are generated from the training halftone images. For architecture-2, i.e., two-level RBF-NN, the first-level RBF-NN is trained with feature vectors tagged with lpi, $k$-level value and gray level. The $k$-level values are generated by down-sampling the original gray images. The gray level tag is not required by the first-level RBF-NN but it is used by the second RBF-NN.

*Evaluation Strategy:* Two methods namely, peak-signal-to-noise ratio (PSNR)and structural similarity index measure (SSIM) [105] often used for measuring image quality are used for judging the efficiency of the proposed inverse halftoning method as well as comparing its performance with some of the well cited previous studies. PSNR value is inversely proportional with the mean square error (MSE). The value of SSIM is computed combining correlation, luminance and contrast. Its range lies in $[0, 1]$. If SSIM be 1 then both images are maximally correlated.

#### 3.2.2.2 Experimental Results and Analysis:

In this sub-section we discussed about result and analysis of art-work based authentication by inverse half-toning methods.

*Inverse Halftone result:* The performance of the inverse transform methods are presented on two pictures namely *Atlas hand* and *Lena eye* which are halftoned at three different resolutions: 60, 70 and 80 lpi. The performance of the architecture-1 is presented first in Table 3.4.

**Table 3.4:** Performance of a single RBF-NN based Inverse Transform Method (Architecture-1)

| Image | LPI | Context Pattern ($m \times m$) | | | | | |
|---|---|---|---|---|---|---|---|
| | | $3 \times 3$ | | $5 \times 5$ | | $7 \times 7$ | |
| | | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| Atlas hand | 60 | 25.052 | 0.859 | ***27.805*** | ***0.901*** | 27.200 | 0.891 |
| | 70 | 24.868 | 0.873 | ***27.283*** | ***0.896*** | 27.240 | 0.890 |
| | 80 | 18.010 | 0.710 | 27.010 | 0.789 | ***28.040*** | ***0.810*** |
| Lena eye | 60 | 21.673 | 0.565 | 18.273 | 0.505 | ***21.797*** | ***0.571*** |
| | 70 | 21.585 | 0.557 | ***23.903*** | ***0.641*** | 23.660 | 0.636 |
| | 80 | 20.735 | 0.514 | ***23.956*** | ***0.624*** | 23.656 | 0.623 |

Note that for the *Atlas hand* PSNR values of around 27-28 and SSIM value of around 0.8 is achieved which is slightly better than what is obtained by the same architecture but without using lpi information [102]. Without using lpi information (i.e. if the training of the neural is done without lpi information) PSNR value of around 24-25 is obtained. The same trend is supported by the *Lena eye* image. This shows lpi information better guides the inverse halftoning process. Context pattern based on which features are extracted around a pixel has definite effect and the result shows that $5 \times 5$ context

produces best average case result. We can further note that similar inverse halftoning methods [102], [103] produced image giving PSNR of around 30-31 when the halftones are based on dispersed dots and printed digitally. The reason is the resolution which is far more (around 150 lpi) for dispersed dot based halftones than it is in line halftones. Next the performance of the two-stage RBF-NN is reported in Table 3.5. Here a $k$-level is approximated image is generated first which is then converted into gray tone. Quality of inverse halftone image has been investigated at different values of $k$. It is to be noted that for all cases, this 2-level RBF-NN architecture gives better performance than single RBF-NN. Though the amount of improvement looks small in terms of PSNR and SSIM values but these improvements are statistically significant for all the three resolutions, $p < 0.05$ by a two-tail t-test. Figures 3.5 and 7.1 show the inverse halftone images at different resolution levels. Results for $5 \times 5$ context pattern are shown in these figures. From the Tables 3.4 and 3.5, one more observation maybe noted. In these tables, SSIM values varies from 0.5 to 0.9 whereas PSNR varies from 18 to 29 (dB). It is observed that when SSIM increases from 0.2 to 0.8, PSNR increases linearly or nearly following a straight line. However, when SSIM rises to 0.8 or higher, PSNR increases rapidly. This observation is supported by the findings in [105].

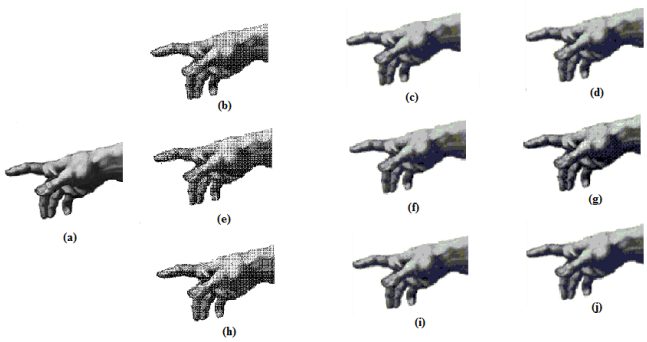



**Figure 3.5:** Inverse Halftoning of *Atlas hand*: (a) original gray image, (b), (e), (h): halftone images at 60, 70 and 80 lpi; (c), (f), (i): inverse halftone images at 60, 70, 80 lpi by using single RBF-NN architecture-1; (d), (g), (j): inverse halftone images at 60, 70, 80 lpi by using two-stage RBF-NN architecture-2.

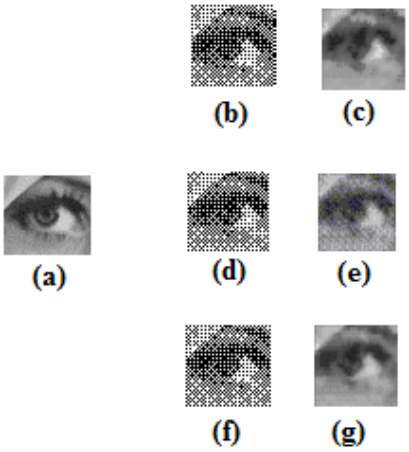



**Figure 3.6:** Inverse Halftoning of *Lena eye*: (a) original gray image, (b), (d), (f): halftone images at 60, 70 and 80 lpi; (c), (e), (g): inverse halftone images at 60, 70, 80 lpi by using two-stage RBF-NN architecture-2.

Next, we compare the performance of our method against the largely cited inverse halftoning methods [106], [107], [108], [109] which have been shown performing well on dispersed halftone images. We checked their performance on line halftone images and results are reported in Table 3.6. The comparison shows that the proposed method outperforms the other methods often by significant margin based on both the metrics, i.e., PSNR and SSIM and for all the three resolutions.

## 3.3 Summary

Criminal activity concerning generation of duplicate copies of certain kinds of security documents has been becoming a potential threat to our civil society. Security artwork authentication is an important issue for this purpose. Fine line or line HT design is commonly using on various type security documents. Authentication or verification is also necessary. This chapter provides an automatic means for verification of authenticity. Here we have proposed two different experiments. First one was microprint level artwork authentication and second one was IHT of line HT document images. The methods

**Table 3.5:** Performance of two-stage RBF-NN based Inverse Transform Method (Architecture-2)

| Image | LPI | $k$ | Context Pattern ($m \times m$) | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | $3 \times 3$ | | $5 \times 5$ | | $7 \times 7$ | |
| | | | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| Atlas hand | 60 | 4 | 24.875 | 0.866 | 27.553 | 0.899 | 27.400 | 0.896 |
| | | 8 | 25.526 | 0.877 | 27.812 | 0.890 | ***27.900*** | ***0.903*** |
| | | 16 | 24.874 | 0.873 | 27.843 | 0.901 | 27.700 | 0.893 |
| | 70 | 4 | 24.884 | 0.870 | 26.778 | 0.889 | 28.705 | 0.891 |
| | | 8 | 25.005 | 0.870 | 26.870 | 0.879 | 27.690 | 0.876 |
| | | 16 | 24.936 | 0.873 | 27.071 | 0.893 | ***29.670*** | ***0.909*** |
| | 80 | 4 | 18.370 | 0.750 | 27.572 | 0.809 | 28.670 | 0.887 |
| | | 8 | 18.887 | 0.768 | 27.670 | 0.890 | ***28.900*** | ***0.894*** |
| | | 16 | 18.910 | 0.784 | 27.550 | 0.886 | 28.400 | 0.865 |
| Lena eye | 60 | 4 | 21.108 | 0.528 | 23.188 | 0.610 | 23.437 | 0.597 |
| | | 8 | 22.875 | 0.582 | 23.753 | 0.613 | ***24.387*** | 0.622 |
| | | 16 | 22.295 | 0.588 | 24.026 | ***0.631*** | 23.921 | 0.619 |
| | 70 | 4 | 21.021 | 0.521 | 23.514 | 0.615 | 24.893 | 0.688 |
| | | 8 | 22.388 | 0.578 | 23.798 | 0.625 | ***26.368*** | ***0.807*** |
| | | 16 | 22.972 | 0.598 | 24.145 | 0.644 | 24.184 | 0.642 |
| | 80 | 4 | 21.219 | 0.518 | 23.735 | 0.601 | 23.904 | 0.616 |
| | | 8 | 22.060 | 0.561 | 25.050 | 0.623 | ***25.066*** | ***0.682*** |
| | | 16 | 22.590 | 0.573 | 24.046 | 0.619 | 24.289 | 0.634 |

**Table 3.6:** Performance Comparison: the first row for each method corresponds to the image *Atlas hand* and the second corresponds to the image *Lena eye*

| IHT Method | 60 LPI | | 70 LPI | | 80 LPI | |
|---|---|---|---|---|---|---|
| | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| LPA-ICI [106] | 25.502 | 0.782 | 26.005 | 0.845 | 26.001 | 0.845 |
| | 24.190 | 0.631 | 24.364 | 0.650 | 24.232 | 0.636 |
| WInHD [107] | 25.589 | 0.862 | 25.596 | 0.862 | 25.592 | 0.862 |
| | 24.306 | 0.655 | 24.722 | 0.687 | 24.370 | 0.657 |
| MAP [108] | 26.996 | 0.892 | 26.999 | 0.892 | 26.995 | 0.892 |
| | 24.136 | 0.621 | 24.455 | 0.634 | 24.170 | 0.617 |
| LUT [109] | 24.885 | 0.780 | 25.213 | 0.781 | 25.211 | 0.780 |
| | 23.341 | 0.621 | 23.806 | 0.630 | 23.428 | 0.624 |
| Our Method | 27.900 | 0.903 | 29.670 | 0.909 | 28.900 | 0.894 |
| | 24.387 | 0.631 | 26.368 | 0.857 | 25.066 | 0.682 |

are essentially based on principles of image processing and pattern recognition. The proposed framework considers design of an efficient but low-cost solution to this problem of detecting duplicate documents.

# Chapter 4

# Printing Technique Verification

*Note: The work of this chapter is related under publications [8].*

## 4.1 Introduction

Printing technique is an impotant security aspect of any security documents. In this chapter, we are mainly concerned about currency note authentication by printing technique verification. Various type printing techniques implies on currency note design like intaglio, non-intaglio printing method. Related security features are importantly contributing for authentication. On chapter 2 and 1 under section 1.3 and 2.3.1, we have seen lots of different security features is contributed for document authentication. But, the process used to print document like banknotes provides important checkup for authentication of the notes. In many cases counterfeiting have been reported even on the paper identical to one as used for genuine notes leaving a very narrow gap to identify the original from the fake. However, the printing technique that is hard to replicate because some of its inherent characteristics. There are numerous printing processes like offset, dry offset, intaglio, letterpress, serigraphy, screen printing, Photostat copying, inkjet, bubble-jet, digital printing, etc. that can be used for printing currency notes. Out of these many possibilities, only a few processes are normally used in practice. For

example, in case of Indian currency notes, dry offset, intaglio, and electronic monitored number printing process are mainly involved. Different printing processes are used to print different parts of a note. However, not all of these printing techniques are applied at a time. The sequence by which the printing processes are executed one after another is itself a security aspect.

Particularly, the final effect of intaglio printing is less bright and less sharp edge as compared to that of offset printing. Examination of final printing effect is an important aspect for verification of security documents' authenticity. Most of these effects are visualized through microscope with high magnification or by different scanning technologies. Forensic experts often take their decision by checking these effects of printing on currency note with mostly the help of a microscope.

After reviewing the existing literature (under section no. 1.3), it is evident that both the geometric and the gray or color level features contribute significantly towards printer identification. However, there is a gap in integrating these two types of features in detecting printing techniques. Moreover, many of the studies illustrated certain geometric properties that would play important role in printer identification but a complete framework starting from computation of features till a decision made by a machine has not been well addressed in many of those studies. Another general shortcoming in the existing studies is the use of synthetic data. The authors generate print outs at lab and then test their algorithms on these samples. Therefore, behavior of these algorithms on real forensic samples is yet to explore. On the other hand, the central goal of this research is to be used for forensic purposes and therefore, forensic community would obviously be interested to know the results when real data is involved.

All these shortcomings motivate us to take up the present research. In this chapter, we attempt to formulate a general framework for authentication of the printing techniques in banknotes.

## 4.2 Authentication of printing technique

The entire approach is based on scrutinizing the printing technique. We consider three different aspects like geometric properties, gray-level features, and color properties for characterizing a particular printer. Most importantly, we attempt to closely follow the practice of the questioned document examiners in detecting printers and try to simulate the same in a machine-vision based framework. Most of the aspects that the forensic experts look for identifying the expected printing technique are computationally grabbed and a machine is configured to give the decision about the authentication of the printing process. This framework is then used for a practical problem namely, identification of fake banknotes based on authentication of printing technique. Here,we involve machine that closely follows a similar approach in order to authenticate the printing process in a bank note. The experiment involves real samples of genuine and fake notes. Results are computed and analyzed to bring out the potential of the proposed framework. The rest of the section discusses about the computation of features and implementation of the proposed method.

### 4.2.1 Features Extraction

Feature extraction in this experiment is largely dominated by the input from the forensic experts. Altogether nine features are extracted which can be broadly classified into three as (i) graylevel features (ii) color features and (iii) structural or geometric features. The features and rationales behind choosing them are explained below.

*Dominant intensity* ($f_p^1$):It is defined by the intensity level that majority of the pixels in the character stroke possess. As the dominant intensity of an image is typical to its process of printing, we use it as a feature. Measuring of this feature requires construction of a suitable mask in order to eliminate most of the background pixels keeping only the

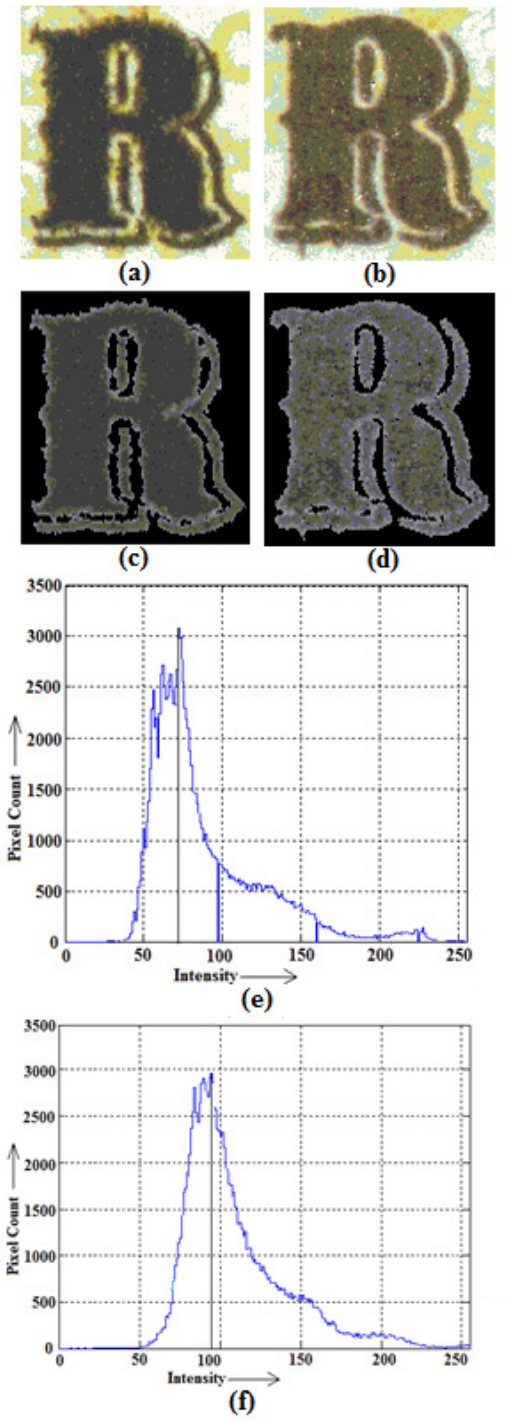



**Figure 4.1:** Role of dominant intensity

character parts. Mathematically this is represented as follows,

$$f_p^1 = x : f(x) = \max(\text{intensity histogram}) \tag{4.1}$$

Figure 4.1 (c) and (d) show the masked images of two character images extracted from two currency notes (one genuine and one fake). Figures 4.1 (e) and (f) show the histograms of gray levels as computed on the masked images.

*Hole count* ($f_p^2$):Number of holes appearing on character strokes gives a significant clue

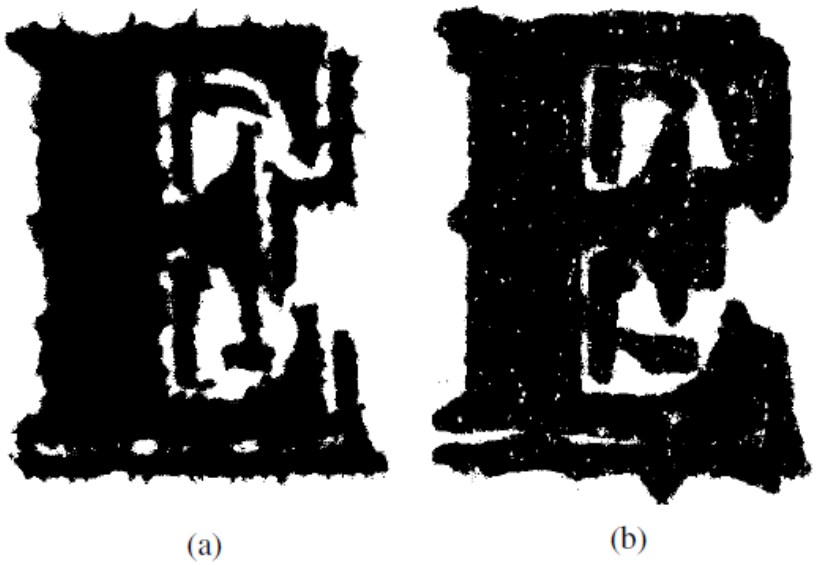



**Figure 4.2:** Holes in acharecter images: (a) a genuine and (b) a fraudulent samples

about the genuineness of a bank note. In binary images of the characters, holes appear as patches of white on the black background of the character. Hole refers to an eight connected white pixel cluster appearing on the character stroke. The ratio of the number of holes to the character area (total area covered by the character stroke) is considered a feature. Mathematically, this is represented as follows -

$$f_p^1 = x : f(x) = \max(\text{intensity histogram}) \tag{4.2}$$

The images in figures 4.2 (a) and (b) clearly (visually) show that this ratio is significantly greater in characters corresponding to fake currencies than in letters of authentic banknotes. Such ratios for images in Figure 4.1 (a) and 4.2 (a) are 0.0009 and 0.0011, respectively, whereas for the images in Figure 4.1 (b) and 4.2 (b) are 0.0015 and 0.0021.

*Average hue* ($f_p^3$): Fake banknotes may sometime appear same as genuine notes in color but by computing the average hue from the character strokes, we may be able to decide whether they are actually printed by using the same technique. This feature is computed in HSV space on the Hue ($H$) stream as follows,

$$f_p^3 = \text{Average}(H) \tag{4.3}$$

Figure 4.3 shows the discriminatory power of this feature for the genuine and fake samples corresponding to figures 4.1 (a) and (b).

*R.M.S. Contrast* ($f_p^4$): The human eye normally fails to capture slight difference in brightness (or glossiness) of two banknotes and this aspect is tactfully used by the counterfeiters. The difference in brightness (or the reflectivity of light) between two samples can be used to detect the method of printing. We capture this feature by computing the RMS contrast [110] of a character images extracted from a banknote. The RMS contrast does not depend on the spatial frequency content or the spatial distribution of intensity in the image. Mathematically it is expressed as follows (where $I_i$ and $\bar{I}$ denote the intensity of the $i$-th pixel and the mean intensity, respectively),

$$f_p^4 = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(I_i - \bar{I})^2}. \tag{4.4}$$

Background masking is not done in this case as the background does not affect the results significantly and thus saves computational time. For example, when the genuine samples as in figures 4.1 (a) and 4.2 (a) are processed, we get 68.7374 and 66.7003 as the values of this feature, whereas the corresponding values are 58.7816 and 55.9164 for

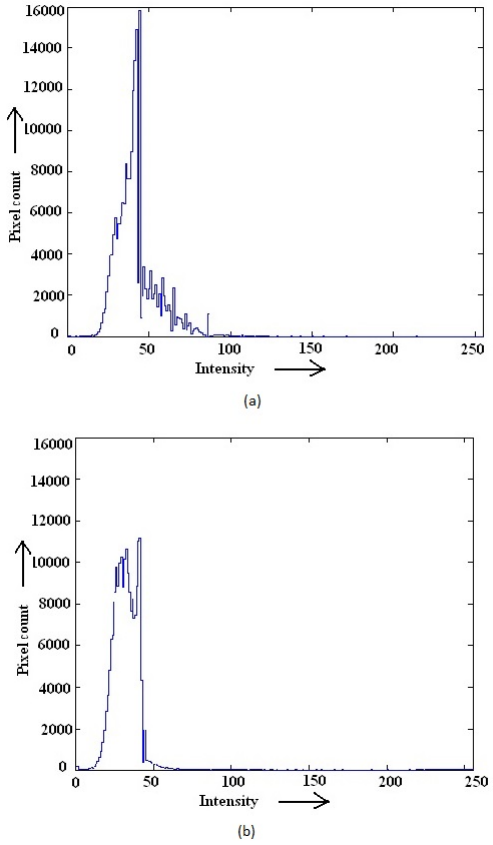



**Figure 4.3:** Histogram ofhue of charecter strokes: (a) genuine and (b) fake samples

the fake samples as shown in figures 4.1 (b) and 4.2 (b).

*Key tone*($f_p^5$): The tonal range of an image refers to its general distribution of intensity. Key tone of an image is represented by the mean gray value of all the pixels. The value of key tone indicates whether the bulk of information in an image is stored in the high/middle/low intensity zone. The value of key tone indeed varies from genuine to fake currency notes due to the difference in pigments used for the printing process. Therefore, we use key tone as a feature in identifying printing techniques.Mathematically, this is represented as follows -

$$\text{Image} \rightarrow \text{Masked character} \rightarrow f_p^5 = \text{Mean}(Intensity) \tag{4.5}$$

*Average Color*($f_p^6$): The amount of a particular color used for printing is different for different printing processes. Average color assesses a reconstituted color matrix, based on a scalar parameter ($p$). Images are transformed from RGB color space to CYMK. While doing this, we observe that in the complemented color space black streams have low values in general. However, input from domain experts suggests that black is one of the principle pigments used in Itaglio printing. So to improve the separability of the notes in the feature space we set a higher ($p$) for the black stream. The principal streams in Intaglio character stroke are blue and black in 500 denominations. The average color is computed as below,

$$f_p^6 = \frac{\sum s(i)}{N} \tag{4.6}$$

where $s(i)$ is defined as

$$s(i) = pB_{\text{blue}}(i) + (1-p)B_{\text{black}}(i), 0 < p < 0.5 \tag{4.7}$$

and $B_{\text{blue}}$ and $B_{\text{black}}$ correspond to blue and black strokes, respectively. Along with these six features, three other features are extracted: edge roughness $E_{PBER}$, ($f_p^7$) (Eq. 4.8), area difference ($f_p^8$) (Eq. 4.9) and correlation coefficient ($f_p^9$). These features are computed based on the work of Breuel et al. [42].

*Edge roughness*($f_p^7$): The interaction of the ink with the substrate (paper) leaves some typical characteristic of the printing process. This issued has been discussed in the papers [4, 42] from where we borrow the following three features that explicitly capture this aspect in order to give a measure to gaze the printing process. The first one is to measure edge roughness. The difference in the amount of smoothening of the character image with respect to its original image on application of an averaging filter is useful to categorize a particular printing process. Here we have used a median filter to smoothen the image. Then the images are converted to binary level using the Otsu threshold. The difference of perimeter of the two images (smoothened image and original image) is

expressed as a ratio as follows:

$$f_p^7 = E_{\text{PBER}} = (p_a - p_b)/p_b \tag{4.8}$$

Here $p_a$ is the perimeter of the actual image, $p_b$ is perimeter of the filtered binary image and $E_{\text{PBER}}$ is the perimeter based edge roughness [42] based on a relative difference of boundary perimeters.

*Area difference* ($f_p^8$) The feature related to area difference [16] is calculated as follows. At first, a character image is binarized using Otsu threshold value. The same image is again binarized using a different threshold value that is calculated by adding a normalized parameter sc to T. The difference in character areas that results on binarization is then expressed as a ratio of the area of the original Otsu-given image as given below. The area difference is computed as -

$$f_p^8 = \text{Area Difference} = \frac{|A_{\text{otsu+sc}} - A_{\text{otsu}}|}{A_{\text{otsu}}}. \tag{4.9}$$

Where, the character image is first binarized using Otsu threshold value (say, $T$) ($A_{\text{otsu}}$) and then the same image is again binarized using a different threshold value that is calculated by adding a normalized parameter sc to T ($A_{\text{otsu+sc}}$).

*Correlation coefficient* ($f_p^9$): Measuring the correlation coefficient between original gray-scale image and the corresponding binary image characterizes individual printerŠs behavior in producing letter contours [4]. This is calculated by using edge images of the gray and binary images. The correlation coefficient is computed as-

$$f_p^9 = \frac{\sum_{(i,j)\in \text{ROI}}(A(i,j) - \bar{A})(B(i,j) - \bar{B})}{\sqrt{\sum_{(i,j)}(A(i,j) - \bar{A})^2}\sqrt{\sum_{(i,j)}(B(i,j) - \bar{B})^2}} \tag{4.10}$$

where $(i,j) \in ROI$, $A$ is the original gray value image, $B$ is the corresponding binary image, $\bar{A}$ and $\bar{B}$ are the mean of $A$ and $B$, respectively.

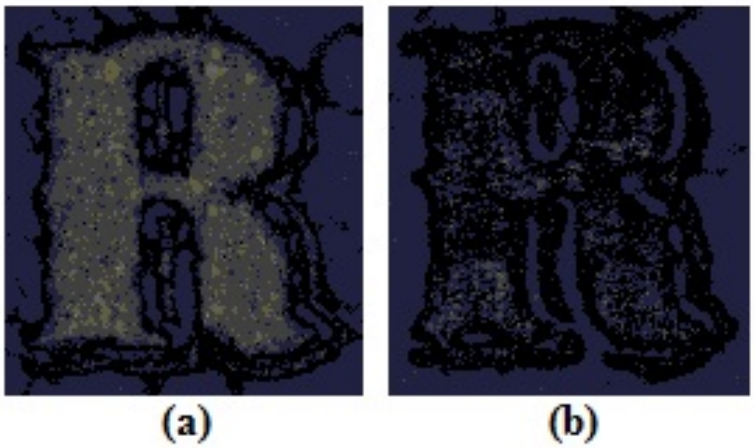



**Figure 4.4:** Parametric color transformation: (a) genuine and (b) fake samples

### 4.2.2 Proposed method

As mentioned earlier that intaglio printing technique is used to print letters and numbers on currency notes. For example, texts like 'RESERVE BANK OF INDIA' or currency note denominations (e.g. 500 or 1000, etc.) are printed using this technique. Authentication of this printing technique is modeled as a 2-class classification problem, i.e. whether the printing technique is the particular intaglio category that is supposed to be used (say, this class is termed as genuine, G) or not (the class representing fake or duplicate, D). Let m be the number of text samples (i.e. character images) known as genuine and n be the number of samples known as duplicate. In the feature space, it is expected that these m samples would form a cluster (CG) and n duplicate samples would form another cluster, CD. To check whether these two clusters are linearly separable, we implement a K-means algorithm and cluster the m+n labeled samples into two classes. Selecting two samples randomly initializes the centers in K-means algorithm. Since K-means results get affected by this initialization phase, K-means is executed more than once (three

times) and each time clustering results are investigated. This investigation reveals that the clusters always overlap and therefore, it is difficult to find a linear decision boundary. Next, support vector machines (SVM) are used aiming at determining the location of decision boundaries that produce the optimal separation of classes. Two types of common non-linear kernel functions namely, polynomial and radial basis function (RBF) are considered. The whole sample set consisting of genuine as well as duplicate samples is divided into four subsets. A fourfold test is conducted so that each subset appears at least once as in training, validation and testing. The proportion in which samples appear in training, validation and test data is 2:1:1 (training: 50%, validation: 25% and testing: 25%).

The classification accuracy is also checked with a Neural Network (NN)-based classifier. An MLP (Multi-Layer Perceptrons) consisting of 9 input nodes correspond to nine dimensions of a feature vector is used. The output consists of only one node to gives binary output (genuine or duplicate). Hidden layer, in the present experiment, contains 2 nodes. A logistic function as explained in the next section is used as the activation function of the network. Like SVM-based classifier a four-fold test is conducted for NN-based classification. Samples appear in training, validation and test data following the ratio 2:1:1.

The final decision about whether the printing technique of a currency note is genuine does not depend on checking of a single character image. As there are many character images on a note, therefore, printing technique is authenticated for number of character images all of which should pass the authenticity criteria. Failure for one image mark the banknote questioned. The decision making process is intentionally made very stringent to reduce false acceptance rate to almost zero.

Finally, a LDA (linear discriminant analysis) is also implemented. The features used in this experiment to authenticate printing technique vary in their power of discrimination. Hence, individual feature wise discrimination power is also studied. Next, features are

sorted based on their decreasing power of classification and then gradually combined to achieve more classification accuracy. All of these methods discussed on 2.3.2.1 section under chapter 2.

**Table 4.1:** Clustering of Currency Note Printing Techniques using K-Means

| Itaration | Distribute of Sample in Clusters | | | | Clustering accuracy (%) =(g1+d2)/2 |
|---|---|---|---|---|---|
| | # Samples in genuine (G) | | Samples in duplicate (D) | | |
| | G (g1) | D (d1) | D (d2) | G (g2) | |
| 1 | 95 | 9 | 91 | 5 | 93% |
| 2 | 93 | 11 | 89 | 7 | 91% |
| 3 | 95 | 7 | 93 | 5 | 94% |
| Avg | 94.3 | 9 | 91 | 5.7 | 92.7% |

## 4.3 Experiments

Magnified scan digitized images of genuine and fake currency notes (Indian rupees of denomination 500) are collected. For considering our study, we consider 100 genuine samples and another 100 samples of fake currency note images. Note that here samples are marked as genuine and fake just based on the fact that whether expected intaglio printing has been used or not to print the samples.

### 4.3.1 Result of k-mean

All the samples are at first clustered using some unsupervised clustering method. The purpose of this clustering is to analyze the distribution of samples in the feature space. The K-means algorithm is used for this purpose. The algorithm finds two clusters one corresponding to genuine samples and another for duplicate samples, i.e. value of K is set to 2. Initialization is done by choosing two samples randomly as to initialize two cluster centers.

The K-means results are evaluated by computing the number similar samples grouped together vs. the number of dissimilar samples contained in that group. Since all samples

are tagged with their classes (genuine or fake) evaluating clustering results in this way is straightforward. Table 4.1 presents the evaluation of Kmeans results. Since cluster centers are initialized randomly, Kmeans were executed three times to get an average result.

### 4.3.2 Classification of SVM

Support vector machines are designed using two different types of non-linear kernel functions namely polynomial and radial basis function (RBF) which followed by equation no 2.2 and 2.3 under 2.3.2.1 section.

The set of 200 samples are divided into four sets to realize a fourfold experiment. In each run of an experiment, two sets are considered as training sets, the remaining two sets serve as validation and test sets. Four different runs were executed. Selecting sets in such a way that each set appears once as a test set and as a validation set (in another run) ensuring that each set would eventually appears twice as training set. The result of this four-fold experiment is reported in Table 4.2 (given on the last page of this paper). It is to be noted that the following values were estimated for the kernel parameters. For polynomial: d = 3 and for RBF: $\gamma$ =1 and they do not change with changing of training sets.

Table-4.2 shows some important observations. For the present problem, polynomial kernel function and RBF based kernel function perform similarly; both the kernels achieve very high classification accuracies. Moreover, the number of support vectors used by the polynomial kernel is far less than the number used by the RBF based kernel. Average number of iterations and norms of weight vectors of polynomial kernel function are far less than those of the RBF kernel function. Mean squared errors (MSE) show that the RBF kernel gives very low MSE when compared to the polynomial (poly) kernel. The MSE is computed by equation no. 2.4 under 2.3.2.1 section.

**Table 4.2:** Classification of Currency Note Printing Techniques Using SVM

| | Support Vector | | Iterations | | Weight Vector | | Accuracy | | MSE | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Poly | RBF | Poly | RBF | Poly | RBF | Poly | RBF | Poly | RBF |
| Run-1 | 3 | 25 | 11 | 16 | 1.0597 | 4.0832 | 100 | 100 | 1.297 | 0.129 |
| Run-2 | 6 | 24 | 4 | 10 | 1.1435 | 4.188 | 100 | 99 | 1,576 | 0.132 |
| Run-3 | 4 | 23 | 3 | 10 | 0.8056 | 4.1843 | 100 | 99.5 | 1.044 | 0.154 |
| Run-4 | 4 | 23 | 6 | 13 | 0.9311 | 4.28 | 99.5 | 100 | 1.354 | 0.112 |
| Avg. | 4.25 | 24 | 6 | 12.3 | 0.9852 | 4.1839 | 99.9 | 99.6 | 1.318 | |

### 4.3.3 Classification of Neural Network

A multilayer perceptron (MLP) is used to design a Neural Network-based classifier. Well-known back propagation algorithm is used to train the network. The network does use of the logistic function as transfer or activation function which mention by equation no. 2.6. And gradient descent method is used to find the optimized set of connection weights that are updated as per equation no. 2.8 under 2.3.2.1 section. In the present experiment, a is set to 0.9 and b is assigned 0.1. The same dataset as used for the SVMbased classifier is also used here to train and test the MLP. Like the SVM-based classification a four-fold experiment is conducted. Experiment shows that like SVM-based classifier NN-based classification too achieves very high accuracy (about 99.5 %, 0.5 % error is attributed to true negative) in classifying printing process as genuine or fake.

**Table 4.3:** Classification of Currency Note Printing Techniques Using LDA

| | Bias(b) | Separability | error |
|---|---|---|---|
| Run-1 | -19.2192 | 5.7692 | 0.5 |
| Run-2 | 0.9741 | 3.3903 | 0 |
| Run-3 | -28.8222 | 7.7076 | 0 |
| Run-4 | -3.7869 | 3.2253 | 0 |
| Avg | -12.71355 | 5.0231 | 0.125 |

### 4.3.4 Classification using LDA

Finally, we use Fisher linear discrimination analysis (LDA) for authentication of printing process. Fisher linear discrimination analysis (LDA) for authentication can derive by different scutioal steps which discussed on section 2.3.2.2.
Like SVM and NN a four-fold experiment is conducted and results are reported in Table 4.3.The average weight vector $\overrightarrow{\omega}$ on maximizing class separablility criterion is 5.0231 whereas the bias b is computed as - 12.7136.

### 4.3.5 Gradation of the features

In this experiment, we use nine features but all of them do not contribute equally in authenticating the printing technique. Figure 4.5 shows their power of discrimination when LDA is used for classification. The correlation feature (i.e. $f_p^9$ as discussed in section-4.2.1) shows the highest discriminatory power (93%) for this purpose. Next is dominant intensity i.e. feature $f_p^1$ (91.1%). The blue line shows the features in their decreasing power of classification. The brown line shows the effect of combining features. When $f_p^9$and $f_p^1$ are combined, the classification rate goes up to 95.4% and finally combination of all the nine features gives an accuracy of about 99.8%.

## 4.4 Summary

This chapter presents a novel experiment on authenticating the printing technique insecurity documents. Printing technique is itself a security feature in most of the security paper documents. Most valuable security document, currency notes take a reference for this experiment. Fraudulent currency notes often could not match the genuine printing technique while producing fake notes. The research embodied in this paper nicely shows that using the standard image analysis and pattern classification techniques an automatic method can be designed to capture many fraudulent cases. Therefore, this

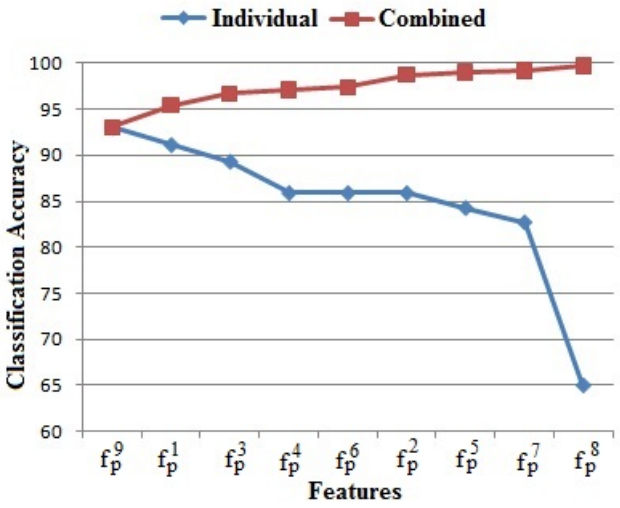



**Figure 4.5:** Parametric color transformation: (a) genuine and (b) fake samples

research provides a viable framework for machine authentication of security documents through verifying the printing technique.

# Chapter 5

# Security Paper Authentication

*Note: The work of this chapter is related under publications [11], [12] and [14].*

## 5.1 Introduction

On Previous work chapter 3 and 4, it was seen that earlier work attempted to exploit several security features but the paper material itself was not consult for automatic authenticating of security document. For any security paper documents including the banknotes, the paper itself plays a crucial role in proving some kind of security to the document. The paper based security is normally achieved by embedding certain special ingredients to the paper material during its manufacturing process. A review on security papers can be found in [18]. Color optical pulp (or fiber) embedded in the paper is an example. Security fibers may be metallic or photo-chromic. The optical pulp defines a certain kind of characteristics of the paper. They are luminescent under ultraviolet (UV) ray and therefore, visible when the paper is scanned under UV or illuminated light ray. The forensic experts often check the intended paper quality by physical contacts and sometimes, though manually, they check the brightness, illumination and density of the paper pulp in order to authentic the paper of the document in question. This section of study attempts to make this process automatic.

The salient contribution of this work is to capture the pulp-based paper security feature in a computational way and then associate these features with the notion of genuine and fake documents. The problem has been viewed from the pattern recognition and artificial intelligence principles. Security aspects are represented as feature vectors and the concept of genuine and fake is defined in the feature space. For extracting features, ideas from rice grain detection [16] in images are borrowed as it closely matches with the present problem of detecting fluorescent paper pulps in images. Next, the features are extracted and analyzed. Classification is done using neural network. Experiment considers Indian banknotes that make use of pulp based paper as a major security aspect. Involvement of real forensic samples is a significant aspect of this study. The experiment shows the importance of paper pulp in detecting fraudulent documents and attests the proposed approach for authenticating banknotes.

## 5.2 Proposed method

The paper used for printing currency notes is a high quality paper made by 100% cotton. Cotton has given whiteness of paper and folding capability. This paper also gives specific identity by its surface finish and crackling sound. During manufacturing process extra features like watermark, security thread and optical fibre (i.e. pulp) are embedded for additional security aspects. The optical fibres or pulps are of specific color and length. For example, in Indian 500 rupee currency note, these fibres are photo-chromic in nature. It spreads randomly on the notes which are illuminated under UV light source. When a banknote is scanned under UV light, the fluorescent paper pulps are visible in the scanned image. Figure 5.1 (a) shows a banknote and Figure 5.1 (b) shows the UV scanned image of the banknote. One may see the fluorescent paper pulps visible in the image in Figure 5.1 (b). The bright spots in the scanned image correspond to the paper pulps present in the note. These pulps play crucial role in authenticating the paper. In a counterfeit note, if the paper is very different from the genuine one, these pulps may

not be seen at all. In a high quality counterfeiting, these pulps came as very bright spots and their shapes show significant difference with respect to the pulp marks of the genuine. Therefore, the illumination and shape of these paper pulps are important in characterizing a note paper as genuine or fake.

Our overall approach is divided into a number of stages: (i) detect pulps in a UV scanned banknote, (ii) extract features from the detected pulps, (iii) train a NN classifier based training samples that include both genuine and fake notes. Overall process depicted on figure 5.2. Once the classifier is trained, we use this for classification which is configured as 2-class (genuine vs. fake) problem.


(a)


(b)

**Figure 5.1:** (a)A 500 rupee Indian banknote (b)UV scanned image of the note

### 5.2.1 Detection of paper pulps

Detection of paper pulps has two stages: identification and verification. During identification phase, detected pulps may be mixed with several foreign (non-pulp) elements

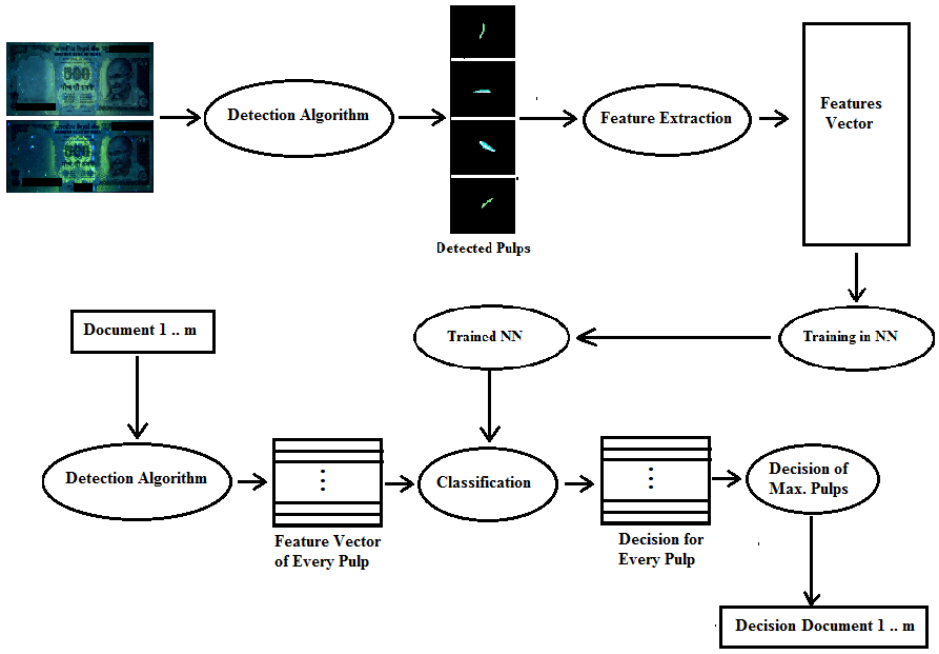



**Figure 5.2:** Diagram of proposed approch

mostly coming from background artworks. So removal of foreign particles is done during verification stage.

**Identification of paper pulps:** Paper pulps are identified in a UV scanned image by following a 7-step method as given in Algorithm- 1. The UV scanned image is represented in RGB color space. As the pulps are mostly blue in color, we convert the RGB image to CMY (Step-2 of the algorithm) and consider the cyan part of the resultant image at Step-3. Next, median filtering is applied at Step-4 to eliminate small unwanted particles. The centroids obtained at Step-7 indicate individual position of pulps in the image. At this stage, all detected points do not correspond to paper pulps. Many other particles which are same as pulp are identified at this stage. These foreign bodies come from background artwork of the banknote. So the next step is to eliminate these foreign elements and identify only the pulps in the image. This elimination is done by the following process.

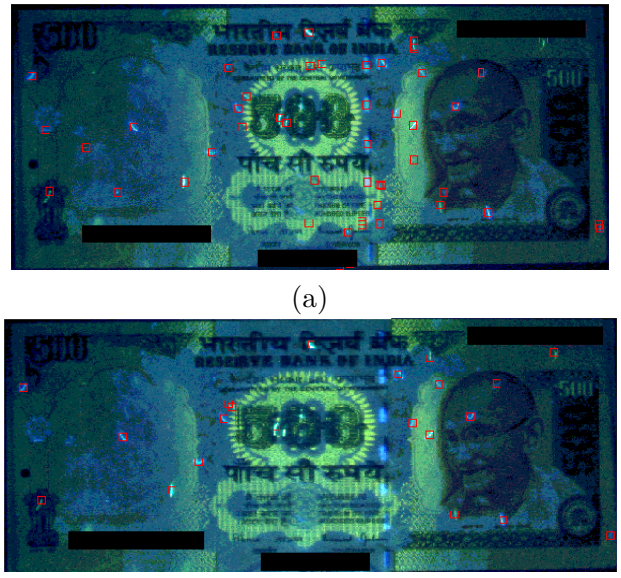
```
1 Begin
2 Step 1: Acquire the currency note image (RGB) by UV light
3 Step 2: Image Complement (RGB − > CMY)
4 Step 3: Extract cyan image
5 Step 4: Apply median filter
6 Step 5: Convert binary image by OTSU thresholding
7 Step 6: Connected component labelling of background pixels
8 Step 7: Compute centroid of each component
9 End
```

**Algorithm 1:** PULP IDENTIFICATION

**Elimination of Non-pulp elements:** The method described in Algorithm-2 eliminates the non-pulp particles from the detected set of pulps. In Algorithm-1, the centroids detected at Step-7 correspond to paper pulps. Here, around each centroid an $m$-by-$m$ pixel-window is considered on the initial RGB image (Step-1 of Algorithm- 2. The value of $m$ is sufficiently large to completely contain a pulp mark within the window. The gray level co-occurrence matrix (GLCM) [111] is computed for each pixel window at Step-2. For this purpose, we transform the gray image to $k$ ($k < 256$) level image ($I$). Let $s \equiv$(x,y) be the position of a pixel in $I$ and $t \equiv (\triangle x, \triangle y)$ be a translation vector. Then the co-occurrence matrix $M_t$ is calculated as,

$$M_t = card(s, s+t) \in R^2 \mid I[s] = i, I[s+t] = j \tag{5.1}$$

Where co-occurrence matrix $M_t$ is a ($k$ x $k$) matrix whose ($i$,$j$)-th element indicates the number of pixel pairs separated by the translation vector $t$ (here, $t = 1$) that have the pair of gray levels ($i$, $j$). Texture features are extracted at Step-3. An artificial neural network (ANN) is used at Step-4 to discriminate pulp from non-pulp elements. A set of training samples is separately identified for the training this ANN. The features extracted at Step-3 are tagged with pulp and non-pulp identification for training the ANN. In our experiment, the values of $m$ and $k$ are set to 60 and 8 (i.e. the image transforms to 8 levels). These values are fitted empirically. Figure 5.3 shows the detection of pulp in Figure 5.3 (a) and then elimination of non-pulp elements to give final result in Figure 5.3 (b). Figure 5.4 (a) shows detection of an individual pulp.



(a)

(b)

**Figure 5.3:** Identification of pulps: (a) detected pulps after execution of Algorithm-1 (b) pulps after elimination of foreign bodies by Algorithm-2

1 **Begin**
2 Step 1: Around each centroid as detected at Step 7 of Algorithm-I, $m \times m$ sub-image is cropped from the initial RGB image.
3 Step 2: For each such sub-image, compute Gray Level Co-occurrence Matrix (GLCM) [111] under consideration of two adjacent pixels on four directions $0°$, $45°$, $90°$,and $135°$ .
4 Step 3: Generate texture level four statistical features i.e. contrast, correlation, energy and homogeneity from each co-occurrence matrix.
5 Step 4: Configure an artificial neural Network (ANN) for discriminating pulps from non-pulp particles.
6 **End**

**Algorithm 2:** ELIMINATION OF NON-PULP ELEMENTS

(a) (b) (c)

**Figure 5.4:** Rectangular box around a pulp: (a) Pulp detection, (b) Region of interest and (c) identification of the rectangular box around the pulp.

### 5.2.2 Feature Extraction from Pulps

Two aspects, namely, shape and color of pulps are considered for feature extraction. Regions of interest are found around the detected pulps. One such example is shown in Figure 5.4 (b). Feature are extracted from this region of interest. Image analysis techniques used for extraction of features. In total, 10 features are extracted: 4 features coming from shape properties and the remaining 6 features are from color properties of the pulp particles. The four shape features are computed as follows:

**(i) Area ($f_1$):** This feature calculates the number of pixels inside pulp identified by a connected component (refer Step 6 of Algorithm-1).

**(ii) Rectangular aspect ratio ($f_2$):** This feature is given by the ratio of the length and width of the rectangular bounding box of the pulp particle. Figure 5.4 (c) shows how the rectangular bounding box of a detected pulp is identified.

**(iii) Pulp Aspect ratio ($f_3$):** The pulp aspect ratio is computed as the ratio of the lengths of the major and minor axes. The length ($d_{max}$) of the major axis is measured as the distance between the end points of the longest line that could be drawn through the pulp particle. Similarly, the length ($d_{min}$) of the minor axis is the distance between the end points of the longest line that could be drawn inside the pulp and is perpendicular to the major axis.

**(iv) Shape factor ($f_4$):** This feature is defined as follows:$f_4=\frac{d_{rms}}{\bar{d}}$; where, $d_{rms}$ is the root means squared deviation and is defined as, $\sqrt{\frac{(d_{max}-\bar{d})^2+(\bar{d}-d_{min})^2}{2}}$. The mean diameter of the pulp is denoted by $\bar{d}$ and computed as $\frac{(d_{max}-d_{min})}{2}$.

**(v) Colour features:**The brightness and illumination of paper pulps give significant clue about the paper quality. They change with the change in paper material. Therefore, features extracted from color space play crucial role in discriminating pulps coming from genuine or fake paper. We consider HSI color space for extracting color features The average Hue, Saturation and Intensity of the pulp pixels give three features $f_5$, $f_6$, and $f_7$. Similarly, their variances are computed and give another three features ($f_8$, $f_9$, and

$f_{10}$).

The above features are considered after consulting with the forensic experts. Many of

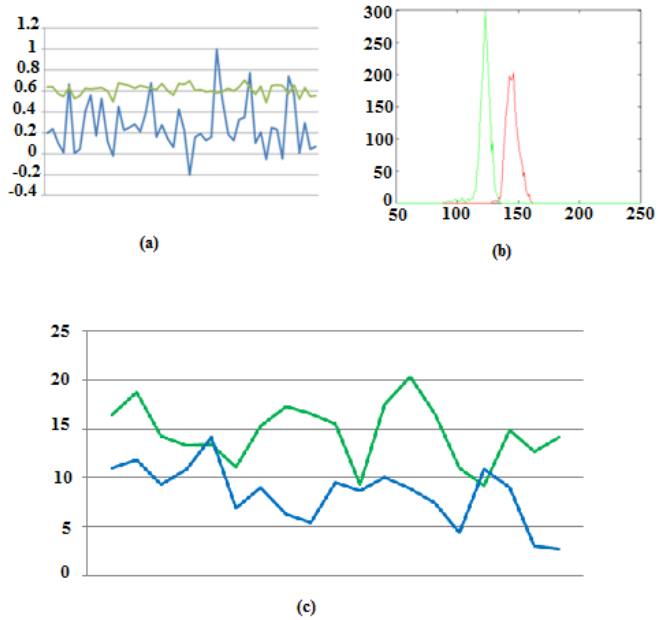



**Figure 5.5:** Discriminatory power of the extracted features: (a) distribution of the pulp aspect ratio ($f_3$) for pulps from genuine and fake banknotes (blue line is for samples from fake currency); (b) distribution of average hue ($f_5$) of pulp pixels coming from genuine (green line) and fake (red line) banknote samples; (c) distribution of the shape factor ($f_4$) for pulps from genuine (green line) and fake banknotes (blue line);

these features they use for manual inspection of the paper in question. It is noted that these features show significant discriminatory power in differentiating genuine and fake samples. This is highlighted in Sec. 3 where experimental results are shown. Figure 5.5 shows the discriminatory power of three features, the first one refers to pulp aspect ratio ($f_3$), the second refers to the average hue ($f_5$) coming from color space analysis and the third, i.e., shape factor ($f_4$) coming from shape analysis.

### 5.2.3 Training of the Classifier

Initially a neural classifier is configured to discriminate whether a pulp is part of genuine or fake paper. A back propagation neural network (BPNN) is used for this purpose.

Multilayer perceptron is used where input layer is consisting of 10 nodes corresponding to 10 features as described in Section 5.2.2. The output layer has just one node as the classification problem is binary in nature. Only one hidden layer is used and the number of nodes in the hidden layer is computed as: N=($\frac{I+O}{2} + \sqrt{y}$);where N=number of nodes in hidden layer; I=number of input features; O=number of outputs; and y=number of patterns in the training set. The multilayer feed forward network model with back propagation (BP) algorithm for training is employed for classification task. A gradient descent method is used to find the optimized set of connection weights that are updated as per equation number 2.8 under 2.3.2.1 section. The efficiency of the BPNN is evaluated using three performance measures i.e. Confusion Matrix, Performance Plot, and ROC plot. The graphical representation of confusion matrix, performance plot and ROC plot in each fold is investigated. The root-mean-square-error (RMSE) is also studied both at the individual pulp and document (i.e. whole paper currency) levels.

### 5.2.4 Authentication of banknotes

Finally, authentication of banknotes is done based on the pulp level authentication. For example, if $p$ number pulps are detected in a UV scanned image of a banknote, each pulp undergoes checking for its authenticity. The neural network described in section 2.3.2.1 is used for this purpose. If majority of the pulps show a particular type (genuine or fake), the banknote turns out of that category.

## 5.3 Experiment

### 5.3.1 Datset

The experiment considers 200 samples of banknotes. All of these are not real samples. We got some real samples from the forensic experts who labelled genuine and fake notes. We extracted features from these labelled notes and labelled the feature vectors as gen-

uine or fake. From these feature vectors, later, we synthetically generated other samples so that we get 100 samples for each genuine and fake classes. We assumed the distribution to be Gaussian to generate the synthetic samples. Each real sample is scanned using *VSC5000* UV scanner. The resolution of scanning was set at 200 dpi. It is noted that each genuine currency note image contains about 15 pulps (this number normally varies from 11 to 17). In fake samples, this number does not vary significantly. In 200 banknotes including both the genuine and fake samples, a total of 3124 pulps were detected. The pulps coming from genuine banknotes are labelled as genuine sample and the pulps originated from the fake banknotes are treated as fake samples. Identification of the pulps above is done following a semi-automatic process. Section 5.2.4 describes a two-stage method for pulp identification. Though the first stage does not require any training, the second stage of this method requires training of a Neural Net. The stage one of the pulp detection algorithm is initially executed for 50 banknotes and extracted pulps are manually tagged as pulp or foreign to train the net. Next, this trained net is used to detect pulps in the remaining 150 notes. It is observed that the net gives about 90% accuracy in discriminating detected pulps as true pulp or foreign element. The errors are then manually corrected to make the dataset suitable for the subsequent experiments. From each pulp, a 10-dimensional feature vector is extracted. Among 3124 feature vectors, 1602 are labelled as genuine and 1522 are tagged as fake. Tagging of each pulp is quite easy as all the pulps extracted from a banknote take the label of that note. The whole dataset is divided into 4 subsets for conducting a four-fold cross validation test. The numbers of samples in training, validation and test sets are in 2:1:1 ratio.

### 5.3.2 Pulp Level Authentication

As mentioned earlier that a neural network is used for discriminating each pulp as genuine or fake. The parameters of the back-propagation neural network are as follows:

**Table 5.1:** Confidence in Pulp Level

| It. No. | Confidence Interval | | Classification of pulps | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Genuine | Fake | Genuine Samples | | | Fake Samples | | |
| | | | G | F | C | F | G | C |
| 1 | (0.975,1.025) | (-0.0150,0.0150) | 47 | 02 | 01 | 45 | 03 | 02 |
| 2 | (0.970,1.029) | (-0.0128,0.0128) | 49 | 01 | 00 | 46 | 01 | 03 |
| 3 | (0.969,1.030) | (-0.0182,0.0182) | 44 | 03 | 03 | 47 | 03 | 00 |
| 4 | (0.970,1.029) | (-0.0250,0.0250) | 48 | 00 | 02 | 43 | 02 | 05 |
| Avg.(%) | | | 94% | 3% | 3% | 90.5% | 4.5% | 5% |

It.: Iteration, G: Genuine, F: Fake, C: Confusion, Accu.: Accuracy

**Table 5.2:** Pulp Level Authentication.

| Fold | #Epoch (Best validation at epoch no.) | MSE (Min) | Gradient | Classification accuracy | |
|---|---|---|---|---|---|
| | | | | Training | Test |
| Fold 1 | 44 (38) | 0.00107300 | 0.00291 | 94.93% | 90.00% |
| Fold 2 | 27 (21) | 0.00784510 | 0.00924 | 92.88% | 88.67% |
| Fold 3 | 27 (21) | 0.07414200 | 0.03070 | 90.91% | 88.11% |
| Fold 4 | 34 (28) | 0.06693800 | 0.04040 | 89.94% | 86.88% |
| Avg. | 33 | 0.03749952 | 0.02081 | 92.16% | 88.41% |

maximum number of epochs: 1000, minimum MSE value: 0.001, learning rate ($\alpha$): 0.9, momentum ($\beta$): 0.1. Two early stopping conditions were used: (a) total mean squared error (MSE)$\leq$ 0.001 (b) training stopped after 1000 epochs. At first, recognition of individual pulps is evaluated without mixing genuine and fake pulps together. In evaluating this, we find out two confidence intervals, one for the genuine pulps and the other for the fake pulps. These two confidence intervals are calculated as $1 \pm [\sigma.Z_{\frac{\alpha}{2}}]$ and $0 \pm [\sigma.Z_{\frac{\alpha}{2}}]$, respectively where $\sigma$ is the standard deviation of pulp recognition accuracy (say, $r$), i.e. $\sigma = \sqrt{(\frac{r.(1-r)}{n})}$, where $n$ is the total number of pulps; $\frac{\alpha}{2}$ represents the area in each of the two tails of the standard normal distribution curve and $Z_{\frac{\alpha}{2}}$ is the two-tailed normal score for the probability of error $\alpha$. Following these confidence intervals, Table-5.1 shows the result for recognition of pulp types at 94% confidence level.

Next all the pulps are mixed together and recognition of their types using the neural net is evaluated. Table- 5.2 reports this result. It is noted that about 88% pulps are accurately classified as genuine or fake by the neural net and this accuracy is achieved at quite low MSE, i.e. 0.037. Fig. 5.6 graphically shows the behaviour of the neural net.

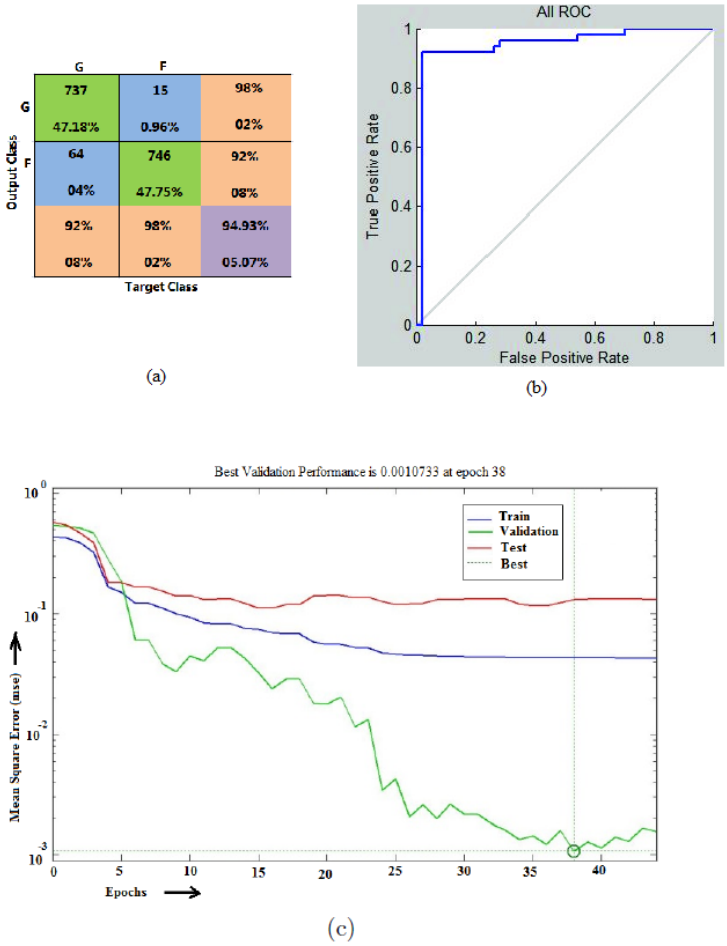



**Figure 5.6:** Behaviour of the neural net in classifying pulps: (a) confusion matrix, (b) ROC plot and (c) performance plot

The results are plotted for fold-1. However, similar characteristic curves were observed for other folds. Fig. 5.6 (a) shows the confusion matrix. The ROC plot is shown in Fig. 5.6 (b). As the ROC plot hugs more the left and top edges, it guarantees better accuracy. Fig. 5.6 (c) shows the performance plotted with mean square error (MSE) value against each epoch. The performance plot shows that with the increase of the number of epoch, the MSE value during training gradually decreases and the best validation is achieved at epoch number 38.

### 5.3.3 Authentication of Banknotes using Pulp Level Authentication

Pulp level authentication result is used to authenticate a banknote as described in section 5.3.2. Let $p$ be the number of pulps detected in the UV scanned image of bank-note after execution of Algorithm- 2 (Section 5.2.1). Each of these $p$ pulps is authenticated using the neural network as reported in Sec. 2.3.2.1. The individual authentication scores are then consulted to determine the nature of the banknote. In Section 5.3.2, we have checked that the classifier can authenticate the paper pulps with about 88% accuracy. Keeping this accuracy in mind, we decided that at least 75% of the pulps in a banknote should be of similar type (genuine or fake) to label the banknote with that type. If it happens that 75% paper pulps do not show agreement in their class label, the system rejects that banknote and calls for manual intervention. For example, if a banknote normally shows 16 paper pulps, at least 12 pulps should have the same class (genuine or fake) for the system to take decision about the category of the banknote.

The above method was tested for authentication of 200 banknotes divided into four (4) groups for conducting a four (4)-fold cross-validation task. The banknotes of the first two folds participated in training of the neural net. Actually, the pulps inside them are used to train the classifier. The third fold is used for validation purpose. The banknotes in the fourth set are authenticated using the trained classifier. It is observed that out of 200 banknotes 199 samples were correctly classified based on their paper pulp. In one case (which is actually a genuine sample), the system fails to decide as some of its paper pulps are degraded because of the degradation of the paper of the banknote. This banknote is an old one and had been folded at many places. For all other cases, 75% or more paper pulps are rightly authenticated for their class and hence, the system could take accurate decision.

Figure 5.7 plots errors in recognizing pulp types as well as document types for eight representative banknotes (genuine samples are marked with green color and red color is used to mark fake samples). The 45° line is shown using the blue color. It is noted

that for each banknote there are some pulps for which types are not recognized properly. Misclassified pulps are marked with '×' for genuine and '+' for fake samples. For these pulps RMSE value is higher. However, as the majority of the pulps in a banknote are classified properly, the overall document level (i.e. at banknote level) RMSE value is quite low. Therefore, banknote note types are correctly identified.

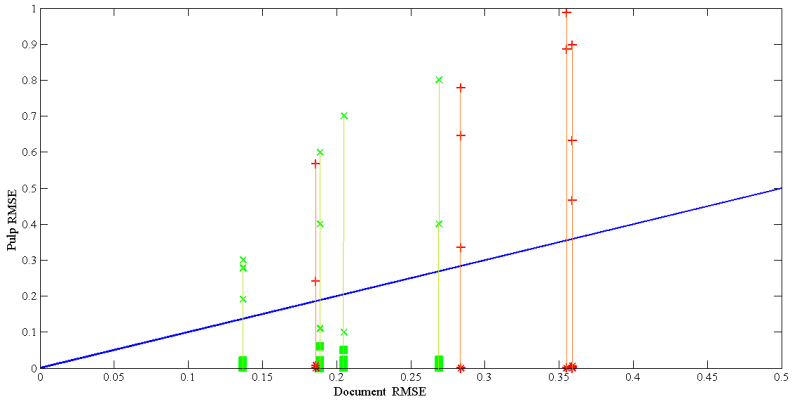



**Figure 5.7:** Banknote classification using paper pulps.

## 5.4 Summary

This chapter reports an interesting experiment in the context of machine aided authentication of security paper documents. The role of paper pulps is investigated in order to authenticate a paper in question. Automatic paper based authentication is fundamental task for security documents like currency note. Experimental-idea endow with rice-grain detection problem. Experiment mainly focused on pulp detection and their authentication proof. Experiments with banknotes strongly attest the viability of the proposed method. Magnified UV scan Indian currency notes used for our experiment.

# Chapter 6

# Ink-Age Analysis

*Note: The work of this chapter is related under publication [7], [9], [10] and journal [3]*

## 6.1 Introduction

Many documents like legal deeds, certificates, university/college mark sheets, bank notes, etc. are common lifelong security documents. Recently, it's seen that forgery of such documents has substantially been increased because of the advancement in printing, scanning, and photocopying technologies. Untrained human eye cannot detect easily these types of fraud documents. Forensic experts try to ascertain the authenticity of these documents through several means [6], one of which is to check the relative or absolute age of a questioned document by determining the age of the printing ink thereon. Therefore, ink age determination has been a standard field of study in forensic science [11]. Ink age determination is a challenging problem because the documents in question may have been generated many decades or even centuries ago. Paper technologies, printing technologies, or chemical component of ink pigment, paper component, etc. all change with time. Moreover, color of ink, shade of colors all differ from one manufacture to other. Also, due to oxidation over a period of time, there would be change in the constituents of chemical substances of the ink used in printing a document.

So far the forensic experts have been following different chemical techniques [1, 24] for determination of ink age. Some of the recently proposed methods can be found in [23,66]. The essence of these techniques is to measure or observe several chemical properties or changes in properties upon use of different chemicals and based on these measures or observations a decision is taken. For case-based experiment these techniques have proved their reliability to an extent. However, many of these methods are destructive in nature and not suitable especially when the document in question has to be preserved as it is. Methods are not fully automated and therefore, they require human intervention. These shortcomings make the existing techniques bit inconvenient especially when a large number of documents in question have to be processed online. Every time a document in question is encountered, if one has to register a case with the police department who in turn ask for help of forensic experts then involvement of so many steps make the entire process delayed and less attractive. Therefore, a fully automated approach for ink age detection would be of great help even to the forensic experts for their own assistance.
This chapter attempts to address this problem from image processing and pattern recognition viewpoint. Only printed documents are considered for the present study. The method is based on the color features of ink used in printing a document in question. The study assumes that (i) there must be entries on the document or similar documents (more specifically it is assumed that similar samples are not varying by more than five years in their ages) whose dates are not questioned that use the same ink formula as the questioned entry or document, and (ii) the questioned and unquestioned ink entries must occur on documents stored under the same conditions to ensure that the entries were exposed to the same environmental conditions. The significant contributions of this chapter include involving color image analysis and pattern recognition techniques for designing automated approach for ink age determination. Finding out the relevant color features and their statistical analysis is also another motivating aspect of this study. The features are simple, easily computable and yet effective for the purpose. Here, we


**Figure 6.1:** (a) LIFE Cover Page (b) LOGO Portion (c) EBONY Cover Page (d) LOGO Portion

also propose a mathematical model on these documents to understand their variation. The main purpose of our model is - 1) to understand the color variation for particular samples OR identify most follower sample on different age segments. 2) Prediction sample's feature after particular time-period 3) testing the proposed model by unknown samples. Nature of this problem is like combinatorial optimization path finding. Ant colony optimization (ACO) [81,112] based method is best fit for that where ants converge towards a short path, expectantly the optimum or a near-optimum solution towards target. The heuristic information can guide the ants toward the most promising solutions,

in a way reminiscent of reinforcement learning. The metaheuristic ACO explore a much larger number of solutions than greedy heuristics. It efficiently solve the combinatorial optimization problems like highly dynamic network routing applications, NP-hard optimization problems etc. Our proposed model has some similarity with this metaheuristic ACO. The types of experiments conducted involving publicly available samples like LIFE (or EBONY) magazine cover pages bring out several exciting observations and open up many new research avenues in the related field.

## 6.2 Scope of Work

The Question for ink age determination or document's life particularly for old printed documents have raised many decades ago or hundred years ago [19, 20]. In our approach we attempt to closely follow the method the forensic experts do in determining ink age. The method is basically analysis of ink colors. The changes of different properties of colors with age are noted. For example, older a document is, much more yellowish and brownish it is. Dark noise also increases with age. The color shade and brightness are affected on a document with its age and this in turn affects hue and saturation level of the ink colors. The main difficulty for these works or may raise question about its ink component, paper and their atmospheric effect. So we have to be cautious about experimental documents or dataset selection. We have followed all documents printed from same manufacturer or printer and long time of gap between document generation and digitization. To avoid the effect of components variation, we concentrated on that period when chemical ink component is not so changing and also on coated cover document considered for less oxidization effect. So, this study is focus on those old documents which are generated in the middle of previous centaury. On that period, component of printing ink are not so changed [113, 114]. To overcome obscurity, every experimented document is selected through its digitized image quality (PSNR) [115], so that extraordinary or unqualified documents can split from our experiment. However, instead of doing

all these analysis by chemically and manually, we attempt to concentrate on statistical analysis of relevant color features that are computationally captured from scanned images. Apart from gray level analysis, we also consider two other color spaces namely, RGB and HSV.

### 6.2.1 Features Extraction

Available digital images of the LIFE magazine cover pages have been considered first in this study. In our approach, we do not consider the entire cover page rather we concentrate on a region appearance of which remains the same in every issue of the magazine. Dealing with the entire cover page is always not effective as color combination changes from one issue to another making relative comparison of different samples difficult. On the other hand, the magazine name (or the logo) is a good candidate for this purpose. In every issue it appears ( Fig 6.1 (a) and (b)) in a rectangular area where text appears in white on a red background. The area of this rectangular region (logo portion) is found to be good enough for the purpose of color analysis. After extracting (done manually in the present study) the region of interest (ROI), it went through a bi-clustering process to separate the foreground from the background. The K-means is applied for bi-clustering and the cluster consisting white pixels (i.e. relative more white than the pixels in other cluster) represents the foreground. This separation is done to study the age effect on the foreground and the background colors separately. Features are extracted from background and foreground separately, analyzed and then the analysis results are combined to take decision. Human experts also do analysis in a similar way, i.e. studying the background and the foreground colors and their changes separately. However, certain color features are of course extracted from the whole region, i.e. considering background and foreground together. Features are extracted from gray, RGB, and HSV color spaces. Initially, many more features than the ones used in this study were extracted and the following 33 real valued features are finally chosen because

of their significant discriminative power as observed on a training set. Choice of these features also involve consultation forensic practitioners,

*Average intensities*: The first set of features includes the average gray levels of the entire ROI, the background pixels, and the foreground pixels. Next, average red, green and hue levels are computed from background and foreground separately. This gives nine (09) feature values (4 on background, 4 on foreground and 1 from the whole region).

*Prominent intensities*: The most prominent intensity levels are computed on four channels namely, gray, blue, hue, and saturation. Prominent intensity level of a channel is computed as the value observed in maximum number of pixels. Prominent gray intensity is computed from the whole ROI and prominent blue, hue and saturation levels are computed from background and foreground pixels separately. This gives seven (07) features, three from background, three from foreground and one from the both.

*Pixel profile*: The most prominent intensities are recorded as explained just now. The numbers of pixels showing these prominent intensities are also noted. These numbers are normalized with respect to total number of pixels as image size may slightly vary from one sample to another. For example, while computing the intensities on background, the numbers of pixels are normalized with respect to the total number of background pixels and similarly for foreground. So this pixel profile adds seven (07) more features.

*Kurtosis of image colors*: Apart from the highest intensity levels, we also measure the kurtosis to analysis whether variations are due to infrequent extreme deviations. Kurtosis is separately measured for blue, hue, and saturation levels on background as well as on foreground. Therefore, we get six (06) additional features for these measures.

*Geometric mean*: It is noted that geometric means of hue and saturation level significantly differ in two pages if they vary in ages. Therefore, we compute these means separately for background and foreground pixels. Four (04) more features are then added to capture this aspect.

## 6.3 Proposed Method

After features are extracted from every sample images, which are spread over five decades, their dating is to be done. That process applied on similar types of document which was printed many decades ago. Color features vector applied linear and nonlinear classifier to determine its decade which discussed on subsection 6.3.1. One of the important questions is still remaining, this study cannot predict future document quality or cannot guide its potentiality for manufacturer, printers or question document examiner. So, it's important to understand the trend of color shade variation rather than its classification rate. Subsection 6.3.2 and Subsection 6.3.3 of our study are focusing in this direction.

### 6.3.1 Document Dating by Classifier

This is modeled as a 5-class pattern recognition problem, i.e. which decade a particular document belongs. Let $m_i$ be the number of samples of $i$-th decade. In the feature space, it is expected that these $m_i$ samples would form a cluster ($C_i$). If any two clusters $C_i$ and $C_j$ are linearly separable then the task of decision-making becomes easier. If $D_i(X)$'s are the linear decision functions, then a given document, $X$ belongs to $C_i$ if $D_i(X) > D_j(X)$, j $\neq$ i. However, a simple investigation reveals that the clusters are not linearly separable. For this purpose, we implement a K-means algorithm and cluster the N $(=\sum_{i=1}^{5}(m_i))$ labeled samples into five classes. Selecting five samples randomly initializes the centers in K-means algorithm. Since K-means results get affected by this initialization phase, K-means is executed more than once (three times) and each time the clustering results are investigated. This investigation reveals that the clusters always overlap and therefore, it is difficult to find $D_i(X)$, i.e. linear decision functions. The classification accuracy is then checked with a Neural Network (NN)-based classifier. An MLP (Multi-Layer Perceptrons) based NN structure used. The MLP consists of 33 input nodes correspond to 33 dimensions of a feature vector. The output consists of five nodes

corresponding to five decades. The hidden layer, in the present experiment, is made of 3 nodes. A Gaussian Radial Basis Function as explained in the next section is used as the activation function of the network. This method depicted on Figure 6.2.

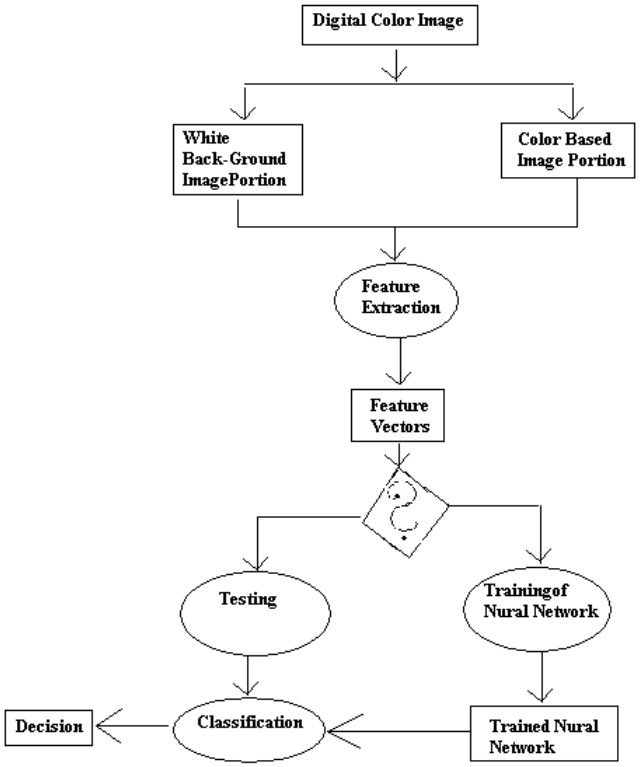



**Figure 6.2:** Flow chart of document dating by Classifier

### 6.3.2 Modelling of Ink-Color Degradation

Next proposed model came from an evolutionary meta-heuristic based algorithm like Ant Colony Optimization (ACO). This algorithm is inspired by ant's social behavior. Ants are no sight to find shortest way from food to their nest. They find with the help of chemical materials. This chemical component called pheromone which is left on the course while moving. They lay pheromone trails on the graph edges and choose their path with respect to pheromone which progressively decrease by evaporation. Ants prefer to

move along those nodes which connected by short edges with a high amount pheromone. In addition, artificial ants have some extra features that do not find their counterpart to real ants. They are usually associated with data structures that contain memory from their previous action. The amount of pheromone deposition and evaporation are done by function of the path quality. The artificial ants choose an edge often been depending not only the pheromone, but also on some specific local heuristics. Directly, ACO is very difficult to incorporate here. The basic differences with our proposal are - On ACO, nodes are not belonging in stations where ants can travel any directions and final path identify by minimum cost but not maximum follower. Figure 6.3 depicts our

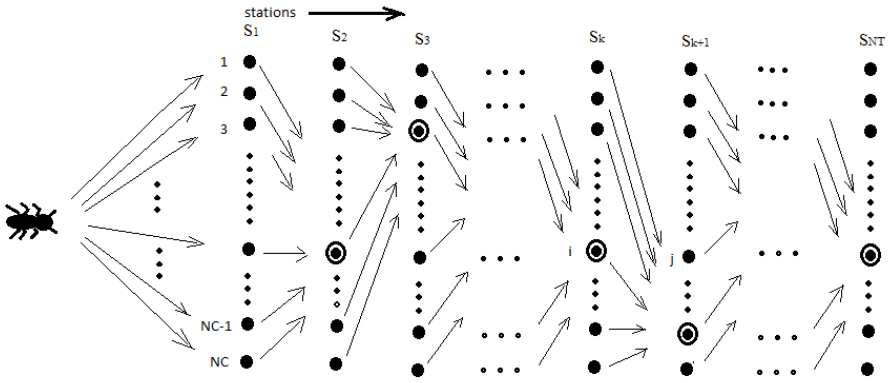



**Figure 6.3:** Proposed model graph

proposed model. Every decade represent by station where number of samples present i.e. called nodes. Suppose, $NT$ numbers of decades are represented as $S_1$, $S_2$,..$S_{NT}$ stations and every station present $NC$ number samples or nodes. Before starting iteration, $N$ numbers ants are randomly placed on $NC$ number nodes. Suppose, in every iteration these $N$ number ants are starting their journey from $S_1$ station and move sequentially toward next station which ultimately end their journy by $S_{NT}$ stations. Every node, an ant selects only one path from $NC$ number of paths which direct next node of next conjugative station. If every path denote by (i,j), then $i$th node belongs $S_k$ station and $j$th node belongs $S_{k+1}$ station. For this path selection, an ant uses a heuristic factor,

denoted by $\eta_{ij}^k$ and pheromone factor, denoted by $\tau_{ij}^k$. These are indicators of how good it seems to have node $j$ at node $i$ of the permutation. The heuristics value is generated by some problem dependent heuristics whereas the pheromone factor stems from former ants that have found good solutions. The next node is chosen by ants from the following probability-

$$P_{ij}^k = \frac{[\eta_{ij}^k]^\beta * [\tau_{ij}^k]^\alpha}{\sum_{j=1}^{NC} [\eta_{ij}^k]^\beta * [\tau_{ij}^k]^\alpha} \tag{6.1}$$

Where, $i$=1,2,3,...,NC $\in k^{th}$ station.

$j$=1,2,3,...,NC $\in (k+1)^{th}$ station.

k=1,2,...,(NT-1); $\alpha$, $\geq$ and $\beta$ are positive constant that determinate the relative importance of pheromone and heuristic value.

For selection of ant's path (i,j) depends on maximum value of $P_{ij}$ from rule (5.1). The initial heuristic value $\eta_{ij}$ is denoted by-

$$\eta_{ij}^k = \frac{Q_1}{dist(S_k(i), S_{k+1}(j))} \tag{6.2}$$

Where, dist$(S_k(i), S_{k+1}(j))$ = Eculidean distance and $Q_1$ be the constant positive value. The initial pheromone value $\tau_{ij}$ is denoted by

$$\tau_{ij}^k = \frac{dist(S_k(i), S_{k+1}(j)) * Q_2}{\sum_{i=1}^{NC} dist(S_k(i), S_{k+1}(j))} \tag{6.3}$$

Where, $Q_2$ be the constant positive value and dist($S_k(i)$,$S_{k+1}(j)$) be the equilibrium distance between two node i and j for two different conjugative station i.e. $S_k$ and $S_{k+1}$. The local pheromone updation for particular path(i,j) is intended by one complete move

of an ant. Here i$^{th}$ station to j$^{th}$ station updation done by following rule-

$$\tau_{ij}^{k} = \left\{ \begin{array}{ll} (1-\rho) * \tau_{ij}^{k} + \rho * \Delta_{ij}^{k} & \text{if } P_{ij}^{k} \text{ max for all } j \text{ w.r.t. } i \\ (1-\rho) * \tau_{ij}^{k} & \text{Otherwise} \end{array} \right\} \tag{6.4}$$

Where $\Delta_{ij}^{k} = \frac{C}{tag^{k}(i) * tag^{(k+1)}(i)}$ and C is the positive constant value and $\rho(\epsilon(0,1))$ is the persistence of trial for particular path (i,j). (1-$\rho$) represented pheromone evaporation none other (i,j) path. This tagging value (i.e. tag $^{k}$(i)) is generated by using K-mean clustering algorithm. Set of individual cluster nodes specified individual tag no. i.e. the nodes class which have individually minimum average distance, is given minimum tag value. Consecutively, the nodes class which have individually maximum average distance is given maximum tag value.

The global pheromone value update after each iteration, i.e. when ants completed their tour from starting station ($S_1$) to ending station ($S_{NT}$). Then, it is computed by following rules-

$$\tau_{ij}^{k} = (1-\rho) * \tau_{ij}^{k} + \rho * \Delta_{ij}^{k} \tag{6.5}$$

Where, $\Delta_{ij}^{k} = \frac{R}{\eta_{j}^{k+1}}$; Here R is constant.

Here, j be the particular node on (k+1)$^{th}$ station which have maximum probability $P_{ij}^{k}$ for (i,j) path and $\eta_{j}^{k+1}$ is represented by maximum number of ants come from $k^{th}$ station to $j^{th}$ node. $\sigma$ be the constant positive real no. where $\sigma\epsilon$ (0,1).

The pheromone-updating rule was meant to simulate the change of amount of pheromone due to both pheromone depositions and evaporations on visited and unvisited edge respectively. This algorithm stops when all ants found a solution by passeing same nodes or completed particular number of iteration.

- ***Input:*** NT,NC, N, $\alpha, \beta, \rho, \sigma$ C, R, $Q_1, Q_2, I_{max}$
- ***output:*** State Best tour node (N), Distance of best tour

**Step 1:** Compute initial heuristic value ($\eta_{ij}^k$) and pheromone value ($\tau_{ij}^k$) for every path (i,j) in between $k^{th}$ station and $(k+1)^{th}$ station by equation 6.2 and 6.3. Where

Station counter k= 1 to NT-1.
$k^{th}$ station node counter i= 1 to NC.
$(k+1)^{th}$ station node counter j= 1 to NC.

**Step 2:** N no of ants randomly placed on NC nodes on starting station $S_1$ (Where station counter k=1). If $N_i$ ants placed on $i^{th}$ node, then $\sum N_i = N$.
**Step 3:** Every ant serch next node on next consecutive station.
**for** *Node Counter i:=1 to NC* **do**
  **if** $N_i > 0$ **then**
    From equation no 6.1 compute probabilty ($P_{ij}^k$) to place $N_i$ no of ants on $(K+1)^{th}$ station.
    **else**
      Continue
    **end**
  **end**
**end**
Repeat until all ants passes to next consecutive station
**Step 4:** Update the local pheromone
**for** *i:=1 to NC from* $k^{th}$ *station* **do**
  **for** *j:=1 to NC from* $(k+1)^{th}$ *station* **do**
    Pheromone update by equation 6.4
  **end**
**end**
**Step 5:** Repeat the step 3 and 4 with k=k+1. Until all ants placed on $S_{NT}$
**Step 6:** Update the global pheromone
**for** *station counter k+1:=NT to 2* **do**
  select node j on k+1th station with max. ($P_{ij}^k$)
  Pheromone update by equation 6.5
**end**
**Step 7:** Check the stop criterion
**if** *(Iteration* $I < I_Max$*) and (*$I_Max$ *Not stagnation behavior)* **then**
  Record the best tour of maximum ants follow,
  Go to Step 2.
  **else**
    Print the node of best tour & distance of best tour
    STOP.
  **end**
**end**

**Algorithm 3:** Ink-Color Degradation

### 6.3.3 Apply Proposed Model on Old Printed Documents

The objective of our proposed model is to select the best node from every station where all or maximum numbers of ants have passed. The design of color level variation is represented in Figure 5.2. In our study, number of station NT considered as five for five-decade samples. Where $S_1$ station is represented by '70 decade samples, similarly, $S_2$, $S_3$,$S_4$,$S_5$ station are represented by '60, '50, '40, '30 decade samples respectively. Every station 40 samples are placed on 40 nodes. So NC is represented by 40. Out of total 200 samples, 40 samples of $S_1$ station are represented as starting-nodes and 40 samples of $S_5$ station are represented as ending-nodes. Every iteration ant starts from one of the nodes from $S_1$ station and ends on one of the nodes from S5 station. Except first station, this model ultimately selects one particular node from every station where all ants or maximum number ants pass through it.

Algorithm- 3 represents the solution for the optimal path design of Ink-Color degradation. Here, the best tour path represents the document of future decade. If any documents start from $S_1$ station then we can guess that it can pass through this path as it become older. So the document which belongs in '70 decade can predict future resentation (i.e. document) on different decade by best tour path.

## 6.4 Experiments Result and Analysis

Old LIFE magazine cover page (http://books.google.com/) have considered first for this experiment which shown on figure- 6.1(a). Historical background and manufacturing periods are helpful for consideration these documents. Life magazine was introduced on 4th January 1883, in a New York City artist's studio [116]. After certain flicker circumstances, Time founder Henry Luce bought this magazine in 1936 [117]. Then, time Inc. published as a weekly photojournalism news magazine. Luce started to publish Life from November 23, 1936 until it has last issue on December 29th 1972. After then LIFE

launch as a weekly newspaper supplement for few years. The website, life.com, stays alive from March 2009 to January 2012. Before this period, Google already began hosting a documentation of the magazine's photographs from November 18, 2008, as a joint effort with Life [118]. All magazines are digitized after so long period of fabricating time. The full digitized documents for that time period (1936-1972) are available through Google Book Search. Here, our experiment has done through these scan document images. Another same type of old magazine i.e. EBONY cover page considered for verification purpose. Available digitized documents on the period of 1970 -2010 considered here. One of this magazine cover pages is depicted on figure 6.1(c).

Both magazine cover pages, scanned at 100dpi, as available with the Google book project ($Googlebooksathttp : //books.google.com/, accessed4thjanuary2014$) have been used here. Age can easily mark from publishing date which specified on magazine cover page. Five different set of samples are collected from five consecutive decades. Every set have six to eight years fabricating time gap. For constructing our proposed model, we have considered first LIFE magazine. First set of samples fitted by'30s i.e. 1936 to 1937. Accordingly fifth one fitted by'70s i.e. 1971 to 1972. Only logo portion 6.1(b) of this cover page considered for our experiment because this reverse red color logo portion have been unchanged and followed same art-work on every issue. Here, this logo is considered as Region of Interest (ROI) for an experimental sample. ROI portion of each sample calculated PSNR value with respect to one standard sample. If the computed PSNR value has up to the mark then that document considered for next stage of the experiment. Finally 40 samples have selected from every decade. In total, 200 samples (40 for each decade) are there in the dataset.

The numerical results are presented in this section. All computation of our proposal is implemented in Matlab 7.0. Results of linear classifier discussed on subsection 6.4.1 and non-linear classifier NN descussed on subsection 6.4.2. On subsection 6.4.3 discussed about result of ink color degradation model.

### 6.4.1 Results of K-Means

**Table 6.1:** K-means Result

| | Distribution of samples in different clusters(Ci) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Iteration | Ci (30's samples) Samples:40 | | Ci (40's samples) Samples:40 | | Ci (50's samples) Samples:40 | | Ci (60's samples) Samples:40 | | Ci (70's samples) Samples:40 | |
| | 30's samples | Others | 40's samples | Others | 50's samples | Others | 60's samples | Others | 70's samples | Others |
| 1 | 17 | 24 | 20 | 15 | 24 | 17 | 27 | 15 | 29 | 12 |
| 2 | 29 | 10 | 29 | 8 | 31 | 13 | 32 | 8 | 35 | 5 |
| 3 | 25 | 13 | 26 | 10 | 29 | 14 | 31 | 10 | 31 | 11 |
| Average | 23.7 | 15.7 | 25 | 11 | 28 | 14.7 | 30 | 11 | 31.7 | 9.3 |
| Accuracy | 59.2 | | 62.5 | | 70 | | 75 | | 79.2 | |

All the 200 samples are at first clustered using an unsupervised clustering method. The purpose of this clustering is to analyze the distribution of samples in the feature space. The K-means algorithm is used for this purpose. Euclidean distance is used to find similarity between two samples. The algorithm finds five clusters corresponding to five decades. The k-means results are evaluated by computing the number similar samples grouped together vs. the number of dissimilar samples contained in that group. Since each sample is tagged with its year of publication evaluating clustering results in this way is straightforward. Table 6.1 presents the evaluation of K-means results. Since cluster centers are initialized randomly, Kmeans were executed three times to get an average result. From Table 6.1, overlapping of samples in the feature space can easily be visualized. This indeed discards designing of a classification system based on any linear decision functions.

Another important observation is that the more old the samples are, less accurate is their clustering, e.g. 30's samples are clustered with 59.2 % accuracy which eventually improves to give an accuracy of 79.2% in clustering 70's samples. Similar trend is also observed in determining age as explained next. We analyzed and found that the reason behind this trend is that the ageing effect on older documents is less consistent than it is in relatively newer samples.

### 6.4.2 Age Determination by Neural Network

An MLP is next designed for determining age of a sample by using a Neural Network (NN)-based classifier. The generalized function of NN discussed on section 2.3.2.1 under equation number 2.5. Where $\phi_j(x)$ is the transfer or activation function that takes form of a Gaussian Radial Basis Function (RBF) as follow under equation number 2.7.
The set of 200 samples are divided into 4 sets to realize a four-fold experiment. The proportion in which samples appear in training, validation and test data is 2:1:1 (i.e. training: 50 %, validation: 25 % and testing: 25 %). Sets are selected in such a way so that each set appears at least once as a test set and a validation set. To ensure that each set would eventually appears twice as training set, four different runs were executed. The result of this four-fold experiment is reported in table 6.2. It shows NN-based classification gives an average accuracy of 74.5% in determining ink age on the test set. Relatively newer samples show better performance than the older samples. The reason is as stated before that the aging process is more random as ink grows more and more old. Table 6.3 shows the confusion matrix in dating the samples (each sample appears as a test sample). Here also, it is interesting to observe that dating errors are not very random in nature. Confusion occurs with samples of the neighboring decades only when the samples are relatively but confusions for older samples spread over beyond the neighboring decades.

### 6.4.3 Result of Ink-Color Degradation Model

The numerical results of our probalistic model are presented in this sub-section. It mention earliar that LIFE magazine considered to generate our proposed model. It is discussed on sub- section 6.4.3.1. After than the model is verified by EBONY magazine which discussed on sub-section 6.4.3.2.

**Table 6.2:** Accuracy for Ink Age determination by the Neural Net

| Runs | Correct ink age dating of test dataset (test samples for each decade=10) | | | | | |
|---|---|---|---|---|---|---|
| | 30's | 40's | 50's | 60's | 70's | Total |
| Run1 | 5 | 7 | 9 | 8 | 8 | 37 |
| Run2 | 6 | 6 | 8 | 7 | 8 | 35 |
| Run3 | 8 | 7 | 6 | 9 | 9 | 39 |
| Run4 | 6 | 7 | 8 | 8 | 9 | 38 |
| Avg. | 6.25 | 6.75 | 7.75 | 8 | 8.5 | 149 (74.5) |

**Table 6.3:** Confusion Matrix for Age determination

| | 30's | 40's | 50's | 60's | 70's |
|---|---|---|---|---|---|
| 30's | 25 | 5 | 6 | 2 | 8 |
| 40's | 5 | 27 | 5 | 3 | |
| 50's | 1 | 4 | 31 | 4 | |
| 60's | | | 3 | 32 | 5 |
| 70's | | | | 6 | 34 |

#### 6.4.3.1 Experiment on LIFE Magazine

In section 2 we have already stated that old LIFE magazine is considered to building experimental model. Total 200 samples are considered for this experiment. Every documents are statistically judged by image quality matrices i.e. PSNR values. PSNR values of all selected documents meet the criteria of 25 or more. Tagging value is also generated from each node which belongs 1 to 8. If minimum tag value consider as one (1) than eight (8) be the maximum tag value. Our model experimented with set of parameters. The best parameters values of present experiment are -

$\alpha$ =1, $\beta$ =4, $\rho$ =0.1, $\sigma$ =0.1, C=10, R=50, Q1=10, Q2=10,$I_{max}$ =200

We have started our experiment with considering 30 no. of ants (N) which achieved un-stagnation stage up to Imax. But, If we appllied N=60, 90 and 120 resepectively then the method achieved stagnation stage before Imax. Respective figure 6.4, figure 6.5, and figure 6.6 are showing the distance of best tour with respect to (w.r.t.) iteration. This best tour distance is the total distance by popular node (max. Or all ants are

passing through) on each station where starting station considers its middle node. From the above figures (figure- 6.4, figure 6.5 and 6.6) it is seen that after certain number of iteration, curve become parallel to x-axis. At this stage the process is in stagnation stage. It is also observed that total tour distance on stagnation stage be less as number of ants are increased. Figure 6.7 represent ants behaviors on stagnation stage. From figure 6.7, it can say that minium no of ants require to achieve stagnation stage. After than, if number of ants be increased then more number of iteration require to achieve stagnation. In stagnation stage maximum or all ants passes through particular nodes. This particular node in every decade represented most follower document on that age-segment. From these selected document color shade eminence can also predict. The quality of this outcome judge on next paragraph. Two ways it can judge the efficiency

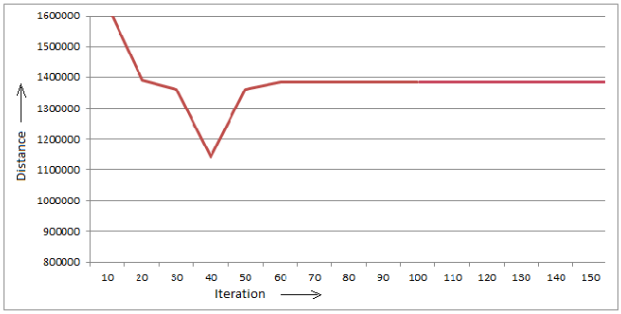



**Figure 6.4:** Distance w.r.t. Iteration when no. of ants is 60

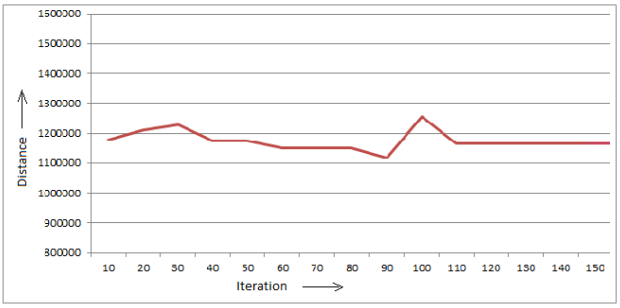



**Figure 6.5:** Distance w.r.t. Iteration when no. of ants is 90

of final outcome. 1) To generate future sample and check with its original one - which is

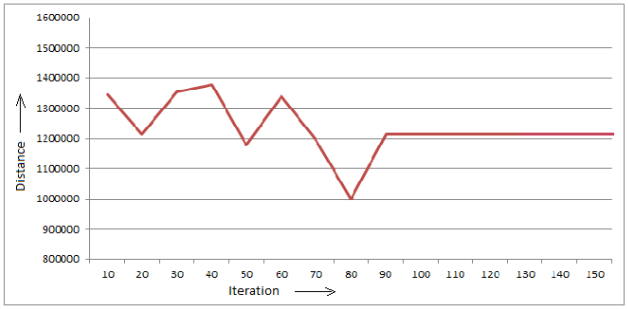



**Figure 6.6:** Distance w.r.t. Iteration when no. of ants is 120

quite difficult task in present conditions. 2) Classification accuracy checking - which is done next. Table 6.4 shows the accuracy rate of our proposed model for N=120 no. of ants. Here 20 samples are considered from each decade for test. Every sample is identifying its station by calculating minimum equilibrium distance from the resultant nodes of every station. This experiment gives 76 % accuracy rates which increased rather than the previous experimental result.

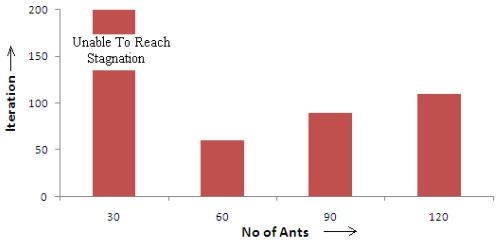



**Figure 6.7:** Stagnation graph according iteration w.r.t. no. of ants

#### 6.4.3.2 Verify by EBONY Magazine

Proposed model are again verified by another digitized old magazine EBONY. EBONY is another old weekly magazine which has been published more than 70 years. Cover page logo portion is considered for our experiment. 6.1 (c) and (d) is depicted the cover page and logo portion respectively. From 1970 to 2010 we have considered five decades samples for investigation. Samples in every decade have eight to ten years gap. Here

**Table 6.4:** Accuracy of proposed model for N=120

| Samples | 70's Decade | 60's Decade | 50's Decade | 40's Decade | 30's Decade | Correct |
|---|---|---|---|---|---|---|
| | S1 | S2 | S3 | S4 | S5 | Accuracy Rate |
| S1 | 12 | 4 | 2 | 2 | 0 | 60 |
| S2 | 2 | 17 | 1 | 0 | 0 | 85 |
| S3 | 0 | 2 | 15 | 3 | 0 | 75 |
| S4 | 0 | 0 | 1 | 17 | 2 | 85 |
| S5 | 0 | 1 | 0 | 4 | 15 | 75 |
| | | | | | Total | 76 |

also 200 documents are selected for experiment by same way. In this case PSNR values of every sample have set on above 25. The parameters are also set with following values- $\alpha$ =1, $\beta$ =3, $\rho$ =0.1, $\sigma$ =0.1, C=11, R=50, Q1=12, Q2=12,$I_{max}$ =200
Here stagnation stage achieved before Imax when number of ants N was 60, 90 and 120. But N=30, the system have not achieved stagnation stage. After constructed the model, it is examined by 20 test samples in each decade. Here, this experiment achieved 72 % accuracy rate which quite similar with previous LIFE magazine experiments.

## 6.5 Summary

In this chapter, we attempt to design ink-Color degradation on old printed paper. Through IP technique, we extracted some selective color-level features which especially affect according ink-age. Firstly pattern recognition principles have provided a viable framework for the purpose of ink age determination. The method is fully automated and shows 74.5 % overall accuracy. For better understating of color-level variation according its age we elaborated on next part of our experiment where these features vector is applied on our proposed optimising algorithm. This algorithm based on designed probabilistic laws with having some flexibility. This approach integrated in probabilistic model, so that it can edify according predicament. LIFE magazine cover page have applied to generate this model and EBONY magazine cover page applied for verification.

On both cases experiment have specified particular sample on different age segment and also shows 76 % and 72 % overall accuracy rate respectively. The overall accuracy results show that the system can assist the human experts with a great extent.

# Chapter 7

# A case Study:Machine Authentication of Indian Currency Note

*Note:* The work of this chapter is related under reffered journal [2].

## 7.1 Introduction

The problem of large scale counterfeiting paper currency poses a serious threat to our society as large amount of fake notes causes economic instability. Counterfeiting of currency notes affects the existence of the monetary equilibrium as its value, velocity, output and welfare may get affected. Most countries that use paper currency for transactions are plagued by this problem. As per the Canadian Government's report [119] Canada has been experiencing a higher concentration of counterfeit banknotes in circulation compared to the G10 countries. Similarly, Kaushal [120] reports that leading economists and thinkers believe the black economy in India could reach 50 per cent of the GDP if the situation is not handled quickly. Reserve Bank of India (RBI) also highlights the alarming rate at which this counterfeits are increasing, the seriousness of the issue, and describes the continuous government efforts to curb this problem. Unfortunately counterfeiters also adapt to the new security features that are incorporated. Criminals continue

to find ways to replicate the currency despite the new banknote security features in place. There have been leaps and bounds in the technical field of counterfeit currencies, and this together with the recent advances in the digital scanning and copying techniques has been an indomitable force.

After reviewing the literature 1.3 we find that there have been very few studies reporting technical and experimental details on how to automatically authenticate currency notes. Many studies rely on one or two features that also can be duplicated using high end technology. As there are so many security features in the currency notes, analyzing only a certain aspect of the note security may not be a good choice. A complete integrated framework has been missing that looks into many aspects like security features in printing, ink, background artwork, watermark, security thread, etc. Another general shortcoming in the existing studies is the use of synthetic data. Many authors generate samples at lab to test their algorithms. Therefore, performances of these algorithms on real forensic samples are yet to be explored.

In chapter 4 we provided conclusive evidence of decision making using printing techniques, but the approach looked only on the printing process and dealt with a small database (i.e. samples of Indian bank notes) to test the initial system performance. Here also real forensic samples, Indian bank notes are taken as a reference. The Indian currency note authentication poses a big technical challenge because of highly heterogeneous concepts involved with Indian currencies. In this chapter, we have incorporated several other security aspects based on ink, security thread and an art work based feature set that supplements the printing style based approach that we discussed in previous chapter 4. Our research is directed to this end which attempts to provide a complete automated approach for detection of counterfeit currency notes. Also a thorough analysis is provided that explores the performance of the embedded security features and their sensitivity. This analysis helps the regulatory bodies understand which security feature are under what kind of threat of breach and what modifications could be done to improve the design, making it less vulnerable to counterfeiting. Detailed experiments were done with real data to support the claim. A comparative study is also reported, involving forensic document experts and bank staff to show the applicability and robustness of the system at a grass root level. The system constructed with popular pattern recognition tools like the k-means, neural network and support vector machines. We have also analysis the robustness of each feature in tackling the problem and suggest sequential order of considering features for maintain maximum accuracy.

## 7.2 Overview of the proposed method

The proposed method is based on image processing [121] and pattern recognition principles [122]. The feature extraction in this experiment is largely dominated by the input from the forensic experts making sure that every aspect of the security features is considered when choosing features. As not all the features used by the experts [123] can be captured computationally, a subset of the features is used. Some new features which are effective for detecting fake notes, but are difficult to check manually, are also added. The feature space is analysed and visualized by using clustering technique. The decision making process is built using two different classifiers: i) Artificial Neural Networks (ANNs) and ii) Support Vector Machines. Furthermore, a Linear Discriminate Analysis is used to measure the performance of each feature. Our feature extraction process considers four different security aspects of the banknote: i) printing technique, ii) ink properties, iii) the thread and (iv) the art work used in designing the note. The features and rationales behind choosing them are explained below.

*Preprocessing:* Different security features are available on different parts of the banknote image. So the initial scanned image needs to be divided into distinct ROIs. The image of the currency note is registered using Hough Transform on a Canny edge detected image (Fig. 7.1(a)). Template matching of the denomination ("500"/"1000") and the Mahatma Gandhi portrait generates two fixed positions as reference around which the rest of the ROIs are extracted (Fig. 7.1(b)). The horizontal strip just above the registered portrait of Gandhi is used to segment the intaglio fonts. Text extraction from this part is done using the vertical and horizontal pixel projection techniques (Fig. 7.2). The biggest black pixel blob in the image is the area to be focussed on for extracting features using latent scan. A vertical strip just beside the denomination is used for the security thread based measures and finally the region between the registered Mahatma Gandhi portrait and the largest black blob is used for the micro-print line features.

## 7.3 Security features extraction from Indian Bank Notes

The security features in a currency note are embraced in various areas which is mention previous chapter under section 2.3.1. In this section of study we are considering mainly on printing techniques, Ink properties, thread and artwork based security features. Features extractions on

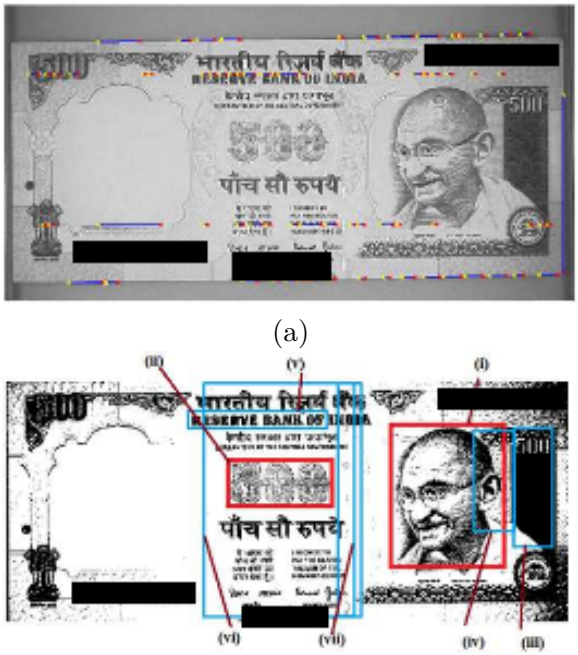



**Figure 7.1:** Preprocessing steps: (a) Image registered using Hough Transform (b) Extracted ROI (i) matched portrait(red)(ii)matched denomination(red) (iii) ROI from latent image(blue) (iv) microprint lines(blue) (v) Intaglio print(blue) (vi) Central pattern(blue) and (vii) Security thread(blue) [bank currency numbers are blackened for security reasons].

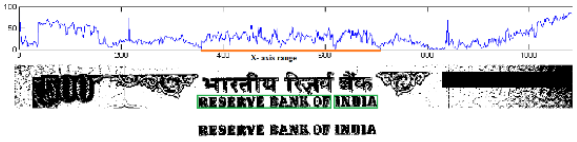



**Figure 7.2:** Text extraction using pixel projection: (top) Black pixel projection on x-axis, (bottom) the horizontal strip used to separate Intaglio fonts in English [bank currency numbers are blackened for security reasons].

these areas are discussed in this section.

### 7.3.1 Printing Technique

Intaglio printing is used for printing currency notes in India. The denomination of the note and RBI (Acronym for the "Reserve Bank of India") are printed on the face of the notes and are always printed using the Intaglio method. This method of printing leaves several signatures that are hard to replicate [28]. We have analyzed some of the features and tried to differentiate

the fake notes from the genuine ones based on printing technique detection. Altogether nine features are extracted which can be broadly classified into three as (i) graylevel features (ii) color features and (iii) structural or geometric features. Dominant intensity ($f_p^1$) and Hole count ($f_p^2$) are graylevel features. Whereas, Average hue ($f_p^3$), R.M.S. Contrast ($f_p^4$), Key tone($f_p^5$) and Average Color($f_p^6$) are color features. And edge roughness ($f_p^7$), area difference ($f_p^8$), correlation coefficient ($f_p^9$) are structural or geometric features. A detailed explanation of how these features were engineered and the reason why they are used can be explored in section 4.2.1 under 4 th chapter. Based on the above nine features, classifiers are trained to identify fake notes based on printing technique.

### 7.3.2 Ink Properties

The reaction of the ink on a particular substrate is different for different inks. This difference actually lends a typical signature indicating the authenticity of a note.

*CCRatio ($f_i^1$)*: Colour composition of the central zone (Fig. 7.3) of a note is analysed by doing an independent component analysis. This was followed by a filtration method that keeps those pixels ON where the green component index in the RGB color space is higher than the blue component index which is in turn higher than the red component index to generate a mask. Computationally it is represented by the color composition ratio ($CCRatio$) feature which is defined as-

$$f_i^1 = \frac{\#\text{ON Pixel in mask}}{\#\text{Pixels in mask}}. \tag{7.1}$$

The number of pixels are fixed as the images are registered using a 4-point registration prior to processing.

*Micro letter ($f_i^2$)*: This feature appears between the vertical band and Mahatma Gandhi portrait in the notes. In notes of denominations 20 and above, the denominational value and "RBI" constitute the micro letters. In our study, we have looked into the colour of these micro letters. The RGB values are first transformed to a specific absolute colour space. This adjustment makes the resulting data device independent. The masked image was changed from RGB to $L^*a^*b^*$ colour space using the CEILAB Illuminant D65 as a reference [124]. The micro letters are printed on a yellow background using a variation of yellow hue. Therefore, a difference of the test note's yellow colour variation against the expected value reveals whether the variance is within specified

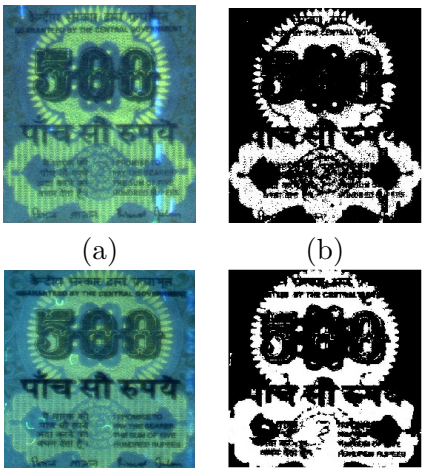


**Figure 7.3:** Analysis of colour composition: **(a)** and **(c)** UV scans of fake and original image; **(b)** and **(d)** are resultant images from the left hand side counterparts after filtration to check colour composition.

limits. In CIE $La^*b^*$ space, +b movement (shift towards yellow is measured on a scale of +b to −b) represents a shift towards yellow along the $b^*$ axis. So we choose $b^*$ stream to study the distribution of yellow on the micro lettering regions. The mean value ($\bar{b^*}$)is computed from 200 genuine random i.i.d. samples. The resultant distribution of the $b^*$ index is plotted in Fig. 7.4. The difference (visually) between genuine and fake notes is seen. Computationally, spread of the index distribution is captured as the feature $f_i^2$ by calculating the standard deviation (spread) values of the $b^*$ index distribution as formulated below,

$$f_i^2 = \text{Spread} = \sqrt{\frac{\sum_{i=1}^{N} b_i^* - \bar{b^*}}{N-1}}. \tag{7.2}$$

*Ink Fluidity ($f_i^3$)*: It is observed that the ink used to print genuine currency notes blot considerably greater than the counterfeit ink. The study of fluidity of ink as a vision based feature was done by Franke et al. [125]. Following this study we developed a feature that would computationally help in decision making about the ink authenticity. Edges of the print were taken and the intensity profile was plotted (Fig. 7.5). We then normalize the curve using an averaging kernel. A steady value is computed in Eq. 7.3 as follows. Let $f(x)$ be the *number of pixels having intensity* $x$, $\forall x \in X = \{x_1, x_2 ... x_n\}$, where $x_1$ refers to the intensity value corresponding to maximum

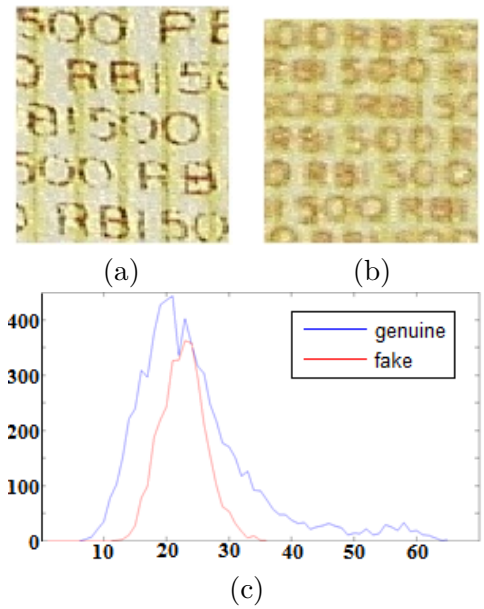



**Figure 7.4:** Analysis of micro lettering: **(a)** genuine banknote, **(b)** fake banknote; **(c)** $b^*$ stream index plot of genuine (blue line) and fake (red line).

pixel count and $x_k = x_{k-1} + 1$. We define $\bigtriangledown f(x)$ as $\bigtriangledown f(x) = f(x_{k+1}) - f(x_k), 0 < k < n-1$ where $x_n > x_{n-1} > x_{n-2} > ... > x_2 > x_1$. $\bigtriangledown f(x)$ computes the differential histogram. This feature reflects the interaction of ink on the substrate, thus checking both the ink and the paper. In other words the printing process. Tampering with either of the two would be reflected in this feature. Fig 7.5 shows visually the difference in edge sharpness. Note that the $\bigtriangledown f(x)$ vector will always start with a negative quantity as the very first value in this vector is the difference from the highest pixel count. The first positive entry is found ($x_p$) and the steady value is generated as follows,

$$steady = \frac{\sum_p^n f(x)}{n + 1 - p} \tag{7.3}$$

where $p$ is the position of the 1st positive entry.The percentage overshoot is then recorded as a feature using the steady value in Eq.(7.4) and computed as follows,

$$f_i^3 = \frac{overshoot}{steady} = \frac{max}{steady} - 1. \tag{7.4}$$

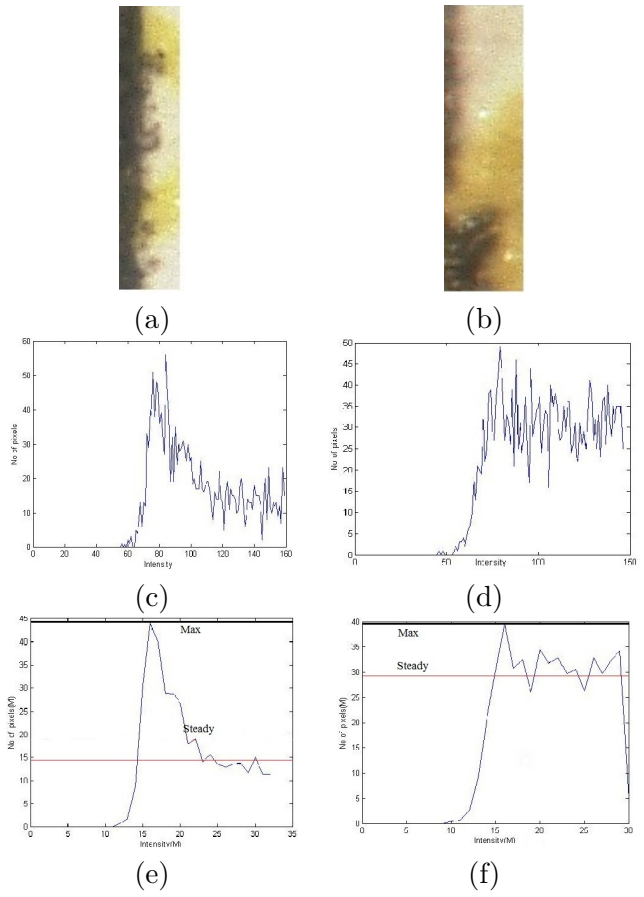



**Figure 7.5:** Ink analysis: **(a)** Genuine ink spread, **(b)** Fake ink spread, **(c)** Histogram of genuine ink, **(d)**, Histogram of fake ink, **(e)** Normalized graph with steady value (genuine ink), and **(f)** Normalized graph with steady value (fake ink).

### 7.3.3 Thread

Two security thread related features are considered: the registration of the notes and the text in the security strip.

*Registration ($f_t^1$)*: The thread should always appear as a single line. This is a way to test of the registration of the notes. We check this using a binary feature, $f_t^1$, which decides whether a note is genuine ($G = 1$) or fake ($D = 0$). Two sets of thick blobs are found (see Fig. 7.6(i)b), one represents the thread parts seen from the front while the other represents the thread parts on the back of the note. Two lines, one for the front and the other for the back, are fit through the centroid ($c_x$,$c_y$) points of the corresponding blobs. The centroid of a blob is calculated using the following formulae where $A$ be the area of the blob. The blobs are treated as non intersecting

polygons defined by $n$ points. Assume C is the set of all centroids, $C = \{c_0, c_1, ..., c_m\}$. Each of the $m$ centroids are defined by a tuple $(c_x, c_y)$. The co-ordinates $c_x$ and $c_y$ are calculated as shown below.

$$c_x = \frac{\sum_{i=0}^{n-1}(x_i + x_{i+1})(x_i y_{i+1} - y_i x_{i+1})}{6A}$$

$$c_y = \frac{\sum_{i=0}^{n-1}(y_i + y_{i+1})(x_i y_{i+1} - y_i x_{i+1})}{6A}$$

This is followed by a distance check of the lines using a threshold distance $t$, empirically computed from 100 note samples. According to the domain experts this is a highly sensitive feature and a good point to do an initial check. We observe that the variance was very low on the order of 0.01 when sampled over 100 genuine images. The expected value from this sampling was set as the threshold $t$.

$$\begin{aligned} d(t_f, t_b) \quad &< t, \quad \text{Genuine,} \\ &>= t, \quad \text{Fake.} \end{aligned} \tag{7.5}$$

where $t_f$ = foreground line points, $t_b$ = background line points. The feature, $f_t^1$ is a binary feature which generates a decision based on Eq. 7.5. Fig. 7.6(i) shows the registration problem of thread in a fake note.

*Text in Thread ($f_t^2$)*: This is another binary feature that checks whether thread text exists. The texts "RBI" and "Bharat" (Hindi for 'India') in Devanagari script are written on original notes where these two words appear alternatively. We extracted the text portion from the threads and then used conventional pattern matching tools to compare. There were only 4 such texts patterns as shown in Fig. 7.6(ii)(a-d) to compare so the templates were extracted from the original image and then used to as ground truth data for pattern matching. A majority of fake notes showed negligible matching because they do not have any text as shown in Fig. 7.6(ii)(f).

### 7.3.4 Art work

This section deals with printing patterns that are intricately introduced in note design. Initially the image is passed through a median ($3 \times 3$ sub window) filter to remove impulsive noise. Next, the centroid of each dot is mapped as shown in Fig. 7.7. The three features described below are extracted and analysed.

*Dot distribution ($f_a^1$)* The distribution of the dot centroids gives us the impression that in

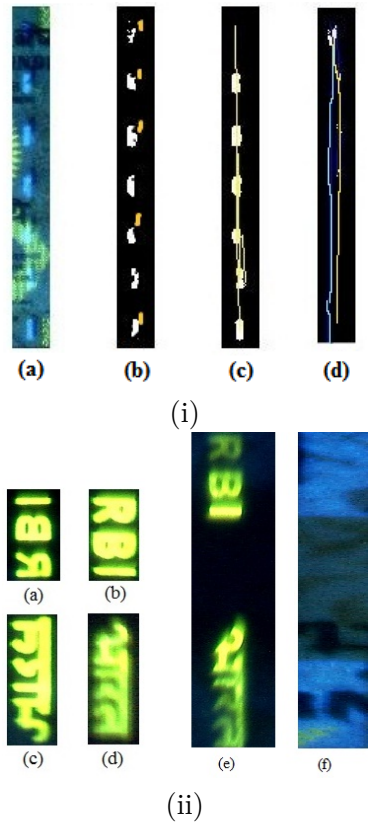



**Figure 7.6:** Analysis of security thread: **(i)** Security Thread: (a) fake note image (b) thick blobs representing the thread on the front (c) line in front (d) two lines which do not overlap; **(ii)** Text in Security Thread: (a)-(d) four occurring patterns (e) original note (f) fake note.

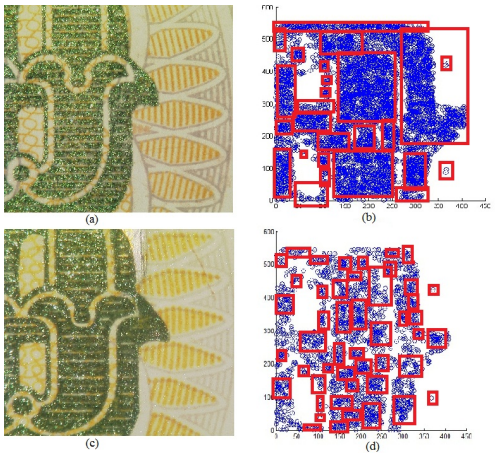



**Figure 7.7:** Analysis of dot distribution: (a) Genuine note (b) distribution of dots centroids for the genuine note (c) Fake note (d) distribution of dot centroids for the fake note.

the fake note the distribution of the dots are far less uniform when compared to the genuine notes. Entropy count provides a measure of this randomness. The entropy ($H$) is calculated as a feature, $f_a^1$, and the following equation measures it as-

$$f_a^1 = H = -\sum_{i=1}^{n} p(x_i) \log p(x_i) \tag{7.6}$$

where $n$ is the number of dots.

*Cluster distribution and dot density features ($f_a^2$ and $f_a^3$)*: We also compute the number of clusters occurring at the character strokes of the letters. An unsupervised agglomerative hierarchical clustering scheme is used with a Euclidean distance check to calculate the number of clusters. Let $C_1$ and $C_2$ be the two clusters and $d$ is the distance matrix, then $max\{d(x, y) : x \in C_1, y \in C_2\}$. An iterative process continues until the separation of the clusters exceeds a threshold (indicated by a Euclidean distance). The cluster density is defined as

$$f_a^2 = \frac{\#\text{Clusters}}{\text{Area of Character Stroke}}. \tag{7.7}$$

From Fig. 7.7 it is evident that the original notes have a more even distribution of dots and far more dots in the character strokes. This difference in the dot count is used as another simple feature which gives the dot density, $f_a^3$ as,

$$f_a^3 = \frac{\#\text{dots}}{\text{Area of Character Stroke}}. \tag{7.8}$$

Adaptive Otsu threshold removes most of the background leaving only the character stroke whose area is calculated.

*Reading Latent Denomination ($f_a^4$)*: When viewing the strip left to the Mahatma Gandhi's picture, the denomination of the note is seen engraved quite distinctively. The machine readability of these denomination digits is measured in Eq. 7.9. This zone comprises of two sets of lines-horizontal and vertical. The vertical lines represent the text part. The sharpness of the lines is different in the two categories of notes (genuine vs. fake) resulting in different readability index $f_a^4$ which is defined as

$$f_a^4 = \text{Score of NN classifier}. \tag{7.9}$$

MLP-NN classifier was trained with digit samples taken from the genuine notes. The vertical lines need to be extracted for the text to be read properly so we use a $3 \times 3$ convolution matrix $(CM)$

$$CM = \begin{bmatrix} -2 & 4 & -2 \\ -2 & 4 & -2 \\ -2 & 4 & -2 \end{bmatrix}$$

that reduces the number of horizontal lines in the image and strengthens the vertical lines. The image that results from this operation has a lot of clutter and noise (Fig. 7.9 (b)). A particular component tag is assigned to each 8-connected cluster and the number of pixels in each cluster is recorded. Fig. 7.9 shows the histogram of component size vs. component tag. A threshold is selected from this histogram to eliminate most of the clutter. This method brings out the latent information from the image. Fig. 7.9 (c) shows the resultant image after thresholding.

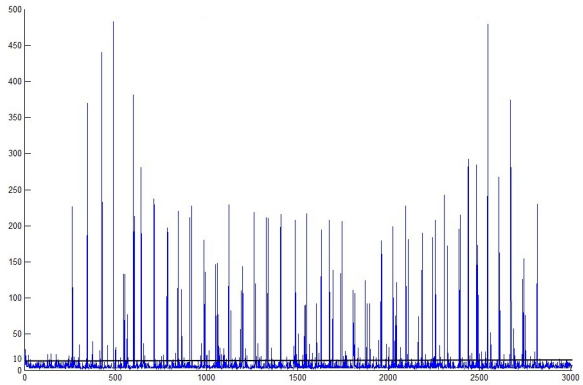
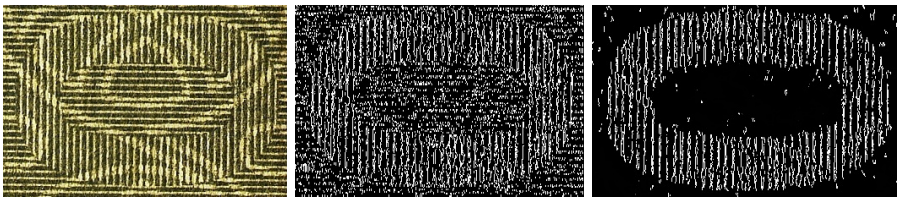



**Figure 7.8:** Component size (along the y-axis) vs. Component tag (along the x-axis).



(a) (b) (c)

**Figure 7.9:** Analysis of latent image: (a) Image of genuine note, (b) after convolution, (c) after filtration. Observe that using only a simple convolution filter doesn't remove all noise, but after filtration the result is substantially better and the digit can be recognised

## 7.4 Experiment

Experiments are conducted to check the validity of the proposed approach using several real life samples. This section explains the data set, the capture conditions, our experimental strategies, and finally, the results.

### 7.4.1 Dataset

The scanning of banknotes involves different medium as explained in the next section. Forensic experts note that a particular group shows a particular kind of proficiency for making fake notes. Two denominations, namely 500 and 1000 (1000 being the highest denomination for Indian bank notes) are considered as the fake note makers mostly target these two denominations. The experts encounter the highest number of fake notes for denomination of 500. In refection to the real scenario, the dataset considers 500 and 1000 denominations in a ratio of 7:3.
The experiment considers 1000 samples (genuine: 500 and fake: 500). All of these are not whole currency notes. Availability of genuine samples is not a problem whereas getting fake note samples is difficult because of security issues. The forensic experts provided us some fake as well as genuine samples. We extracted multiple samples from a single currency note. From Fig. 2.3 one may understand that in a note there are many regions correspond to a particular security feature. For example, samples for intaglio printing can be found in several regions of a banknote. However, regions were selected carefully so that a sample must have all the security features considered in this experiment. Each sample is tagged with its class (genuine or fake).

### 7.4.2 Capturing Conditions

In our experiment, four different scanning techniques: (i) UV Scan, (ii) Latent, (iii) Flood I and (iv) Flood II, are used to image a currency note, using a VSC 5000 (Visual Spectral Comparator) [126]. It uses a spectrometer and other built-in light sources (ultraviolet, infra-red 7.1 on top of visible spectrum light sources) installed specifically to capture images for the task at hand. This type of machine is specially tuned for scanning banknotes, resulting in properties a simple flat bed scanner or a simple photocopier fail to achieve. Anti scan lines over the notes corrupts the output image from simple scanners. These anti scan lines are deliberately incorporated as a design feature, to prevent cheap duplicates from being circulated. An added

advantage of using a specialised high resolution scanner is that it nullifies effects due to ageing or tempering of banknotes and we need not enhance the image as a part of the preprocessing process. Different regions of a note are scanned using different techniques. For example, the central design of the note appears prominent under ultraviolet (UV) light, hence, this region is scanned with UV. There are latent marks on the side of a note. The parameters of the cameras used for imaging are given in Table 7.1. As explained in Section 3, the proposed method captures several features in order to authenticate a note. Not these features are extracted from every image. Four different scans resulting in different features as listed in Table 7.2. It is noted that while scanning different regions of a note under different capture conditions, we have used only one wave length of the light source. This leads to a relatively low cost adaptation of our proposed method as most of the commercial systems and patents (as discussed in Sec. 1) have used high cost multiple light sources of different wavelengths.

**Table 7.1:** Capturing Conditions

| **Condition** | **UV Scan** | **Latent** | **Flood** | |
|---|---|---|---|---|
| | | | I | II |
| Light Source | 365nm UV | Co-axial | 100% | 100% |
| Long pass | VIS | VIS | VIS | VIS |
| Band pass | OFF | OFF | OFF | OFF |
| Magnitude | 2.069 | 9.64 | 29.97 | 40.08 |
| Gain | 12dB | 1 dB | 0 dB | 2 dB |
| Iris | 73% | 50% | 64% | 64% |
| Integration | $\frac{1}{4}$ | $\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{1}{8}$ |

### 7.4.3 Experimental Strategies

The authentication of notes is modeled as a 2-class classification problem: discrimination of genuine (denoted by **G**) vs. fake (denoted by **F**). The reason being that the end users are interested in only knowing whether a note is fake or genuine. Different groups of features are taken into account to accomplish this classification task. The feature groups, as discussed in the previous sections, are based on Ink, Thread, Artwork and Printing techniques. It may be noted that though the problem is considered as 2-class problem, our mail goal is to model the concept of genuineness. Fake samples are considered as negative samples and as we do not assume any prior knowledge about how the fake notes are generated, we consider them homogeneous instead

**Table 7.2:** Source of features

| Features | | Captured from |
|---|---|---|
| Ink | $CC_{ratio}$, $f_i^1$ | UV Scan |
| | Micro letter, $f_i^2$ | Flood I |
| | Ink fluidity, $f_i^3$ | Flood II |
| Thread | Registration, $f_t^1$ | UV Scan |
| | Thread text, $f_t^2$ | UV Scan |
| Artwork | Dot distribution, $f_a^1$ | Flood I |
| | Cluster distribution, $f_a^2$ | Flood I |
| | Dot density, $f_a^3$ | Flood I |
| | Latent denomination, $f_a^4$ | Latent |
| Print | All the nine features related to printing techniques | Flood II |

of heterogeneous samples. Rather, using a prior map of the source to their targeted method of forgery would fail to take into account that these groups are quite dynamic and can change from one method to another.

An initial clustering (bi-clustering) is done on the individual feature groups to visualise their degree of separability in feature space. Classification is done using both Support Vector Machines (SVM) and Artificial Neural Network (ANN) which have already discussed in chapter 2.3.2.1. Four-fold cross validation is followed using both the SVM and ANN based classification. The data set is divided in a 2:1:1 ratio to generate training, validation and test sets. The relative robustness of the individual features is measured by Linear Discriminant Analysis (LDA) to provide a comparative performance analysis of the feature groups. Classification with SVM makes use of two kernels (polynomial and RBF kernel). Mean squared error (MSE) which are computed as on equations 2.2, 2.3, 2.4 under 2.3.2.1 section respectively.

Three different sets of Neural Network classifiers are used for each feature group. Input nodes in each of these networks depend on the number of features i.e. Ink work (3 nodes), Art Work (4 nodes), Printing (9 nodes). Each of the networks have 1 hidden layer and 2 output nodes. The main goal was to realize a simple architecture that would be fast and reliable. In the present form using one hidden layer we got impressive accuracies. Therefore we did not look for higher abstraction of the data using a multi layer perceptron model. It can be argued that that the

query time would not be affected much. However, at the scale of operation which we would operate, Indian economy, small optimizations would result in a huge improvements. The normal back propagation algorithm has been used for training with a sigmoid activation function [equation no 2.6]. A gradient descent method is used to find the optimized set of connection weights that are updated as per equation no 2.8.

Next, we use Fisher linear discrimination analysis (LDA) [127] to study the performance of individual features and their impact on the classification process. The idea is to rank the features based on their degree of separability, being measured by the margin perpendicular to the discriminant hyperplane which elaborately discussed on 2.3.2.1 section. Finally, the cumulative accuracy taking all the features into consideration is computed and a study is conducted to find out a proper sequence in which the features are to be tested. This is important as this reduces the load of the machine if at various stages the number of notes to be checked could be reduced without sacrificing the accuracy.

## 7.5 Experimental Results

At first, results are reported based on individual feature groups. The capacity of these features are then analysed in the context of detecting fake notes. Finally, results are reported considering all the features and a sequence by which features are to be checked is highlighted in order to authenticate a bank note in question.

### 7.5.1 Accuracy of Ink Based Features

There are three ink based features: $f_i^1$,$f_i^2$ and $f_i^3$. The bi-clustering results using these three features are shown in Table 7.3. Table 7.4 reports the classification results when SVM and ANN, respectively are used as classifiers.

### 7.5.2 Accuracy of Art Work Based Feature

There are four features which are captured from the note artwork: $f_a^1$,$f_a^2$, $f_a^3$ and $f_a^4$ as discussed in Section-3. Bi-clustering results using these features are shown in Table 7.5. Classification accuracies by SVM and ANN are reported in Table 7.6.

**Table 7.3:** Bi-Clustering using Ink Based Features

| | #Samples of Genuine (G) | | #Sample of Fake (F) | | Clustering Accuracy |
|---|---|---|---|---|---|
| | G | D | G | D | |
| Iteration1 | 478 | 11 | 22 | 489 | 967 (96.7%) |
| Iteration2 | 471 | 18 | 29 | 482 | 953 (95.3%) |
| Iteration3 | 464 | 22 | 36 | 478 | 942 (94.2%) |
| Avg. | | | | | 95.4% |

**Table 7.4:** Ink Features based classification using SVM and ANN

| | SVM | | | | ANN |
|---|---|---|---|---|---|
| | %Correct | | MSE | | |
| | Poly | RBF | Poly | RBF | |
| Fold1 | 100.0 | 94.0 | 0.491 | 0.530 | 96.00% |
| Fold2 | 98.0 | 94.0 | 0.217 | 0.219 | 100.00% |
| Fold3 | 98.0 | 92.0 | 0.103 | 0.199 | 100.00% |
| Fold4 | 98.0 | 98.0 | 0.064 | 0.256 | 98.00% |
| Avg. | 98.5 | 94.5 | 0.219 | 0.301 | 98.50% |

**Table 7.5:** Bi-Clustering using Artwork Based Features

| | #Samples of Genuine (G) | | #Sample of Fake (F) | | Clustering Accuracy |
|---|---|---|---|---|---|
| | G | D | G | D | |
| Iteration1 | 476 | 15 | 24 | 485 | 961 (96.10%) |
| Iteration2 | 474 | 14 | 26 | 486 | 960 (96.00%) |
| Iteration3 | 478 | 18 | 22 | 482 | 960 (96.00%) |
| Avg. | | | | | 96.0% |

**Table 7.6:** Art work based classification using SVM and ANN

| | SVM | | | | ANN |
|---|---|---|---|---|---|
| | %Correct | | MSE | | |
| | Poly | RBF | Poly | RBF | |
| Fold1 | 99.2 | 98.00 | 0.011721 | 0.00527 | 100.00% |
| Fold2 | 100.0 | 100.00 | 0.018473 | 0.008652 | 99.20% |
| Fold3 | 100.0 | 99.00 | 0.012698 | 0.006285 | 100.00% |
| Fold4 | 98.0 | 99.00 | 0.010526 | 0.008807 | 98.00% |
| Avg. | 99.3 | 99.00 | 0.0133545 | 0.007253 | 99.30% |

### 7.5.3 Accuracy of printing based features

A comprehensive study of the printing based features ($f_p^1$-$f_p^9$) is presented in Table 7.7. In spite

**Table 7.7:** Printing Technique based classification using SVM and ANN

| | SVM | | | | ANN |
|---|---|---|---|---|---|
| | %Correct | | MSE | | |
| | Poly | RBF | Poly | RBF | |
| Fold1 | 100 | 100.00 | 0.000121 | 0.00527 | 100.00% |
| Fold2 | 100 | 100.00 | 0.000873 | 0.008652 | 100% |
| Fold3 | 100 | 100.00 | 0.01028 | 0.006285 | 100% |
| Fold4 | 100 | 98.80 | 0.000205 | 0.008807 | 99.20% |
| Avg. | 100 | 99.75 | 0.002869 | 0.007253 | 99.80% |

of the high accuracy of the individual feature group more than one feature group makes the system more robust. This makes it more immune to the counterfeiter's effort, even if the fake note generators happen to develop means to surpass the security measures of any particular feature set. Use of other features would help in maintaining the high performance of the system. With currency notes we are extra careful in making the system conservative so that no fake currency passes through the system. The two thread based features ($f_t^1$,$f_t^2$) are therefore used upfront that are binary in nature and discriminate notes based on either the presence or absence of two features (namely registration ($f_t^1$) and text($f_2^2$) as defined in Sec.3.2.3) into Genuine (G) and Fake (F). This reduces the processing time too which we have analysed next.

### 7.5.4 Recognition accuracy of the complete system

The performance of individual feature group is quite good but not perfect. Now all the features are taken together to test the system. Individual feature groups are selected in different sequence in order to optimize the performance of the system in terms of accuracy, speed and computational overhead. The Table 7.8 shows that the system operates fastest if the thread based features is placed in the beginning. A total of 1000 samples (500:G, 500:F ) were used to compute the speed of the system. The time reported here comprises the time required to scan the notes using the four scanning methods mentioned in Table 7.1, for registration to figure out which part of the image should be fed to which module and finally to execute the authentication framework. There is a marked improvement in time by using thread based features as the very first step in

processing. However observe that the remaining modules are independent of each other, so to further improve time we parallelize the system as shown in Fig 7.10. A Final decision is taken by using an AND gate on the decision vectors got from the various classifiers. This ensures that whatever passes as genuine has an uncontested majority vote. Even a single negation would result in the note being rejected.

**Table 7.8:** Processing time under different ordering of security features, Genuine (G) and Fake (F)

| Feature Ordering | Test Set | Accuracy | Time (Mins) |
|---|---|---|---|
| 1.Thread<br>2.Ink<br>3.Printing Technique<br>4.Art Work | (F:500<br>G:500) | 100% | 160 |
| 1.Ink<br>2.Thread<br>3.Printing Technique<br>4.Art Work | (F:500<br>G:500) | 100% | 222 |
| 1.Ink<br>2.Printing Technique<br>3.Thread<br>4.Art Work | (F:500<br>G:500) | 100% | 271 |
| 1.Ink<br>2.Printing Technique<br>3.Art Work<br>4.Thread | (F:500<br>G:500) | 100% | 324 |
| 1. Thread<br>Parallel sub modules | (F:500<br>(G:500) | 100% | 64 |

### 7.5.5 Comparison with Respect to Human Experts

We conducted a comparative study to check system's performance with respect to human experts. Two categories of people were involved in this experiment. One was trained forensic experts and the other was bank staffs who are the first in line to receive the bank notes from circulation. The result is tabulated in Table 7.9. At every run the notes were selected at random from Set 2 of the database as mentioned in Sec. 7.4.1

Forensic experts analyze each note individually using various optoelectronic devices which takes quite some time. So, the dataset was kept deliberately small to account for this long process.

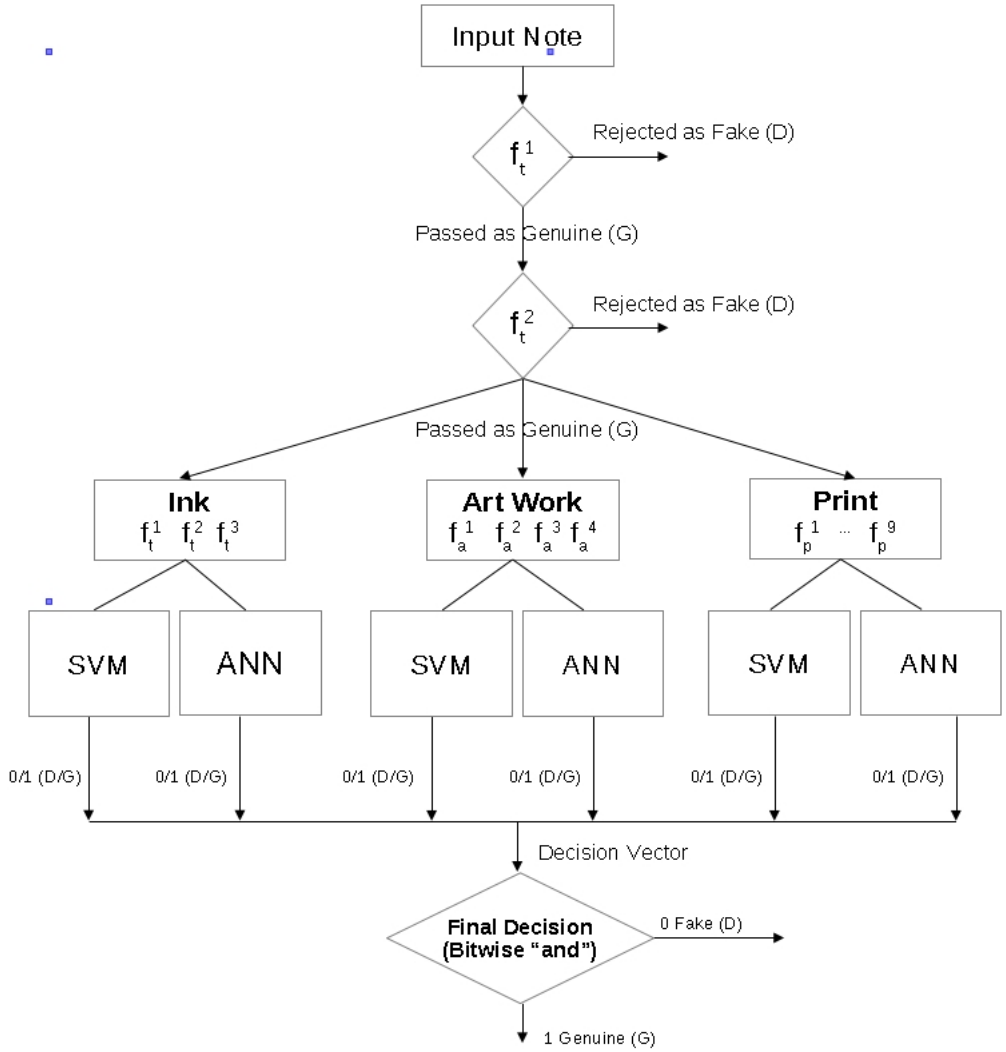
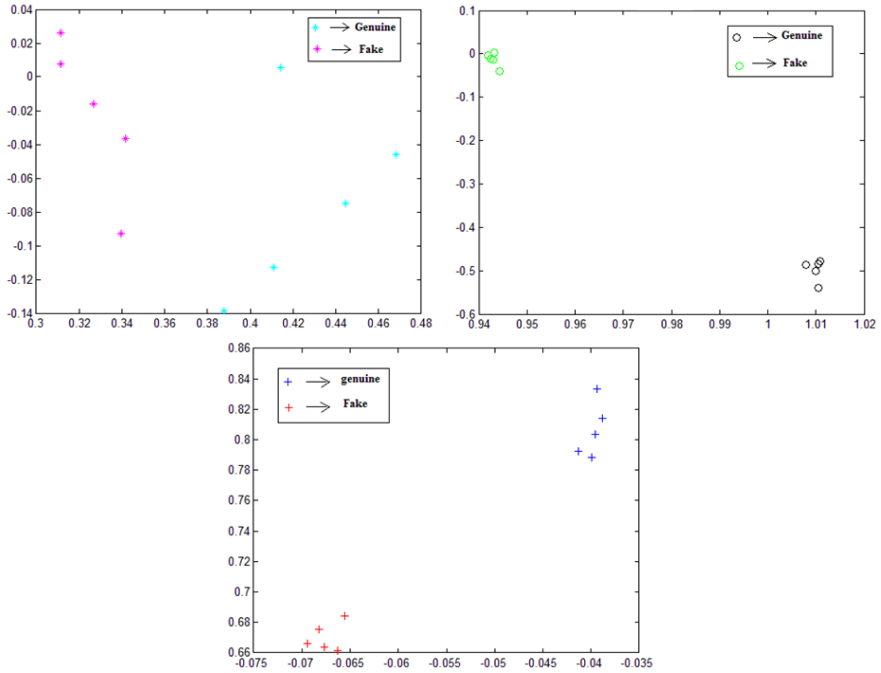



**Figure 7.10:** Work flow of the full system, it starts of with two thread based features deterministically accepting or rejecting notes based on a set threshold. This is followed by feature extraction and prediction from three sub groups of features namely ink, artwork and printing technique using two classifiers. Cumulative decision based on all outcomes gives the final decision.

The results clearly show that the machine outperforms both the forensic experts and the bank staff in accuracy and speed. The bank staff has a reasonable speed of producing results but there accuracy suffers mainly because of the qualitative aspect involved in the authentication process. Bank staffs generally detect notes based on feel of the paper currency. If any suspicious note is encountered they go for further verification but that the lack of sophisticated machines hampers

the overall accuracy. The speed of this system together with its accuracy would be a very good help for the Indian economy that is trying very hard to arrest the counterfeiting efforts.




(a) (b) (c)

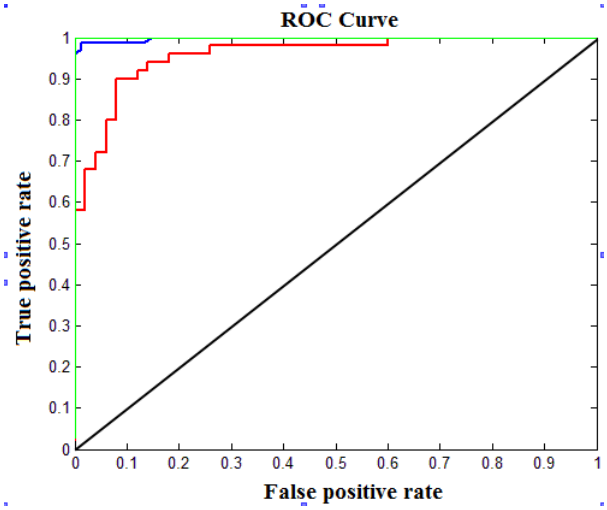
**Figure 7.11:** LDA analysis of the individual feature groups. (a) Ink Features, (b) Art Work and (c) Printing. For all the the cases the problem is casted as a 2 class problem

**Table 7.9:** Performance with respect to Human Subjects

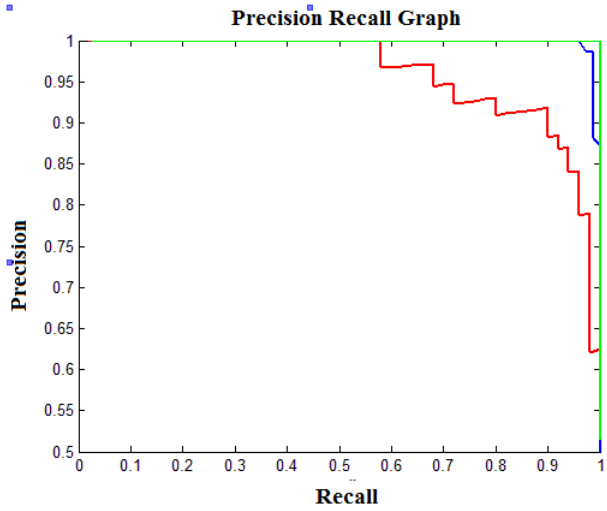
| Testing Method | Sample | Test Result | Accuracy ( Percentage) | Time (mins) |
|---|---|---|---|---|
| Forensic Experts | 50(D) 25(G) | 50(D) 25(G) | 100 | 124 |
| | 50(D) 50(G) | 48(D) 50(G) | 98 | 166 |
| | 25(D) 100(G) | 25(D) 100(G) | 100 | 204 |
| Bank Staff | 50(D) 25(G) | 32(D) 25(G) | 76 | 28 |
| | 50(D) 50(G) | 41(D) 25(G) | 91 | 36 |
| | 25(D) 100(G) | 22(D) 100(G) | 96 | 41 |
| Our System | 50(D) 25(G) | 50(D) 25(G) | 100 | 12 |
| | 50(D) 50(G) | 49(D) 50(G) | 99 | 16 |
| | 25(D) 100(G) | 25(D) 100(G) | 100 | 20 |

### 7.5.6 Relative performance of the feature group

The ability of the three feature groups, namely Ink, Artwork and Printing technique based features, in detecting fake currency notes is analysed using Fisher linear discrimination analysis (LDA) [127]. The previous sub-section 7.4.3 highlights how we use LDA for this purpose. The projection of the individual feature groups are taken on the best discriminant plane and further mapped to show the separability of the feature groups in a 2-D plot. Fig. 7.11 shows the results. The ROC curve shows that printing technique based features are more robust and reliable among




(a)




(b)

**Figure 7.12:** Further analysis is done using : (a) ROC Curve, (b) Recall Precision Curve, the different sub groups of features are Print (green), Ink (red) and Art Work (blue).

the three feature groups. They are followed by art work and finally Ink based features. The steep rise in ROC curves depicts the that the individual feature groups by themselves are extremely robust but not perfect. We also observe that the true positive rate quickly shoots to 1 (ideal) with a very low fraction of false positive. A classifier's performance is judged on the basis of

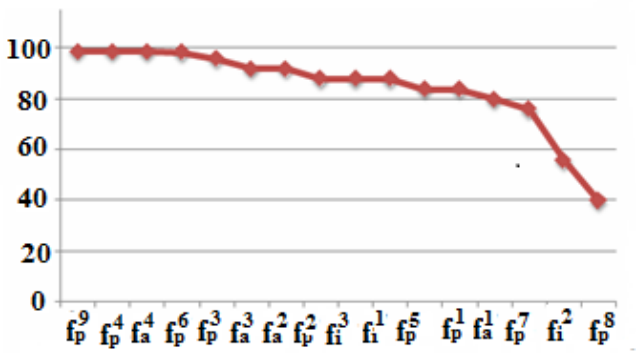



**Figure 7.13:** Performance of individual features. The features are along the $x$-axis and the corresponding accuracies (%) are along the $y$-axis.

how close the precision recall curve Fig. 7.12 is compared to the (1,1) mark which is considered ideal performance of a classifier. We see that all the three feature groups are very close to that mark. Finally, performance of the individual features on a 0-100 % accuracy scale is reported in Fig. 7.11. However, it may be noted that the study on the individual feature strength presented here strongly depends on the samples that are used during the training process and hence, the uniformity of the result presented in Fig. 7.11 is difficult to prove in general. We added this result as the relative performance of the features may give important information to the note designers as to which feature is performing the best and which feature is more susceptible to counterfeiting attack.

## 7.6 Summary

Now-a-days, an automatic method for authentication of currency notes is explored. In this chapter we have experimented on real dataset i.e. Indian bank notes by our proposed framework which followed by different chapters. This study investigates various security features and the method of authentication as followed by the Forensic examiners while dealing with detecting fake paper money. The chapter shows that how many of these features can be computationally

captured in order to automate the authentication process. Several Image Processing and Pattern Recognition methods are exploited to design the overall system. Exhaustive evaluation of the method using real life samples brings out the potential of the approach.

# Chapter 8

# Conclusion

## 8.1 Introduction

In this thesis, authentication of security documents are composed with a complex combination with different security level features which answer some important and common authentication level questions. Throughout history, issuing authorities have faced one common thread i.e. counterfeit of security documents. Today, counterfeiting remains an on-going challenge, influence by the continual technological advancements in reprographic equipment available to both the counterfeiters and general public. The current trends observed in this literature that security documents recognition software is a growing area of research, particularly banknote recognition software is implemented on a wide array of devices from ATMs, banknote sorting machines, self service payment kiosks etc. In literatures it also seen that key computational features can be derived by two broad categories i.e. after capturing image, authentication check by specific ROIs and sensing of intrinsic physical properties associated with a genuine one. Later machine intelligence and speed were integrated for system demand. So, automatic authentication systems rely on a combination of image processing techniques and pattern recognition. In this direction we tried to formulate the answer of security document authentication level question which are generally raised in recent times by forensic peoples. Our research proposals have included with different security features on different types of documents like bank cheques, currency notes, old documents etc.

Our first experiment, Bank cheques were taken as a reference. The proposed method was essentially based on principles of image processing and pattern recognition. Here, we have tested on synthetic data which are generated separately with the help of security design software and others printing devices. First the security features are extracted from the scanned document images and then the notion of 'genuine' vs. 'duplicate' was defined in the feature space. After that Support Vector Machines (SVMs) and Neural Networks (NN) were involved to verify authenticity by these feature space. In this section, a new inverse half toning method has been proposed for reconstructing low resolution line halftone images. This reconstruction is done in order to authenticate an image in question. The reconstructed image is compared with its original image in terms of standard image quality metrics such as peak signal to noise ratio (PSNR) and structural similarity index measure (SSIM). Our proposed inverse halftone technique consists of two parts: at first, the resolution (in lines per inch, lpi) of an input image is estimated and a low level image from the binary line halftone image is constructed. In the second phase, gray level continuous image is generated from the low level description and the lpi information. The method is based on learning based pattern classification techniques namely, neural nets. A comparative study shows that the proposed method outperforms many existing inverse halftone techniques while dealing with line halftone images.

The next challenge of our research was printing technique authentication. Printing technique is itself a security feature in most of the security paper documents. Fraudulent currency notes often could not match the genuine printing technique while producing fake notes. This research embodied in that chapter nicely shows that using the standard image analysis and pattern classification techniques an automatic method have designed to capture many fraudulent cases. Three different classifiers have used in this study i.e. k-means, NN and SVM. This research provides a viable framework for machine authentication of security documents through verifying the printing technique.

Any paper based security document, paper itself is an important security aspect. Next chapter 5 concern with machine aided authentication of security paper documents. The role of paper pulps is investigated in order to authenticate a paper in question. To the best of our knowledge, this was one of pioneering efforts for involving paper pulp for developing automatic authentication of security paper documents. Experiments with banknotes strongly attest the viability of the proposed method. As availability of real forensic data in large scale is a hurdle in every country,

a small set of real forensic samples is used to develop and verify the system.

In the next chapter attempt was made to analys a ink-age on old printed documents. Here also pattern recognition principles have provided a viable framework for the purposed of ink-age determination. Through image processing (IP) technique, we extracted some selective features which especially affect according ink-age. Decision making have done by two different ways. Firstly, classification result was computed by Neural Network (NN) on selective features set. This method was fully automated which previously trained by training dataset. On second attain, one mathematical-model is designed for ink-Color degradation of paper document which is done through image processing (IP) and optimizing technique. The computed feature value is applied on our proposed optimizing algorithm. This algorithm based on designed probabilistic laws with having some flexibility. This approach integrated in probabilistic model, so that it can edify according predicament. In literature, it is a unique and distinctive approach to realize this changing effect.

An automatic method for authentication of currency notes have described in chapter 7. This research is particularly important when the problem of fake bank notes is considered as a serious problem in many countries. Indian bank notes were taken as reference to show how systems have developed for discriminating fake notes from genuine ones. Image processing and pattern recognition techniques are used to design the overall approach. The ability of the embedded security aspects was thoroughly analyzed for detecting fake currencies. Real forensic samples were involved in the experiment that shows a high precision machine have developed for authentication of paper money. This study investigated how the security features can be computationally captured in order to automate the authentication process. Exhaustive evaluation of the method using real life samples were supported the potential of the approach.

## 8.2 Discussions on the outcome of this Research

In chapter 3 presented a pioneering effort to involve machine in checking authenticity of bank cheques. The proposed method first computationally extracts the security features from the document images and then the notion of authenticity vs. duplicity is defined in the feature space. Support Vector Machines (SVMs) and Neural networks (NN) are involved to verify authenticity of these bank cheques. The training, validation and test sets used in different runs of

experiments are exactly the same for designing SVM as well as NN-based classification scheme. Non-linear kernel functions are used to conduct this experiment. Results show that a polynomial kernel based SVM gives about 99% accuracy discriminating duplicate cheques from genuine ones. NN-based classification gives about 97.5% accuracy in classifying test documents as genuine or duplicate. The accuracy is slightly less than that of SVM-based classifier but both the results are definitely comparable. This strongly attests the viability of the proposed approach for machine authentication of printed security documents. In this chapter, a new inverse half toning method has proposed for reconstructing line halftone images. This study presents a neural net based inverse half-toning method for line halftones. The comparison shows that the proposed method outperforms the other existing methods ( i.e. LUT, MAP, LPA-ICI etc.) often by significant margin based on both the metrics, i.e., PSNR and SSIM. As line halftones are generally of low resolution images, inverse halftone images give slightly less PSNR and SSIM value compared to those obtained for dispersed dot (higher resolution) based halftones.

An image analysis based pattern classification method is proposed to authentic the printing process used in printing different texts on currency notes in chapter 4. Features suitable for doing this are selected and then studied to detect fraudulent samples based on the printing method. This classification is also done by using Support Vector Machines and Neural Nets. Polynomial kernel based SVM achieved 99.9% accuracy discriminating result. Whereas, NN-based classification too achieves very high accuracy (about 99.5%, 0.5% error is attributed to true negative) in classifying printing process as genuine or fake. The discriminatory power of the selected features in authenticating the printing process is tested using the Linear Discriminate Analysis (LDA). Like SVM and NN a four-fold experiment, maximizing class separablility criterion is computed 5.0231 and average error showed 0.125% in LDA . Experimental results show that the proposed framework provides a highly accurate framework for authenticating the printing process in bank notes.

In chapter 5 magnified UV scan Indian currency notes used for paper pulps investigation experiment. Using the shape and color features, a multilayer back propagation neural network is used to discriminate paper pulps as genuine or fake. Here, pulp level recognition achieved 94% confidence level. When all the pulps are mixed together to recognition of their types using the neural net then about 88% pulps are accurately classified as genuine or fake. This accuracy is achieved at quite low MSE, i.e. 0.037. Experiment shows that consideration of paper pulps is

quite efficient in authenticating the currency notes.

In chapter 6, we attempt to design ink-Color degradation on old printed paper. Through IP technique, we extracted some selective color-level features which especially affect according ink-age. Firstly pattern recognition principles have provided a viable framework for the purpose of ink age determination. The method is fully automated and shows 74.5% overall accuracy. For better understating of color-level variation according its age we elaborated on next part of our experiment where these features vector is applied on our proposed optimising algorithm. This algorithm based on designed probabilistic laws with having some flexibility. This approach integrated in probabilistic model, so that it can edify according predicament. LIFE magazine cover page have applied to generate this model and EBONY magazine cover page applied for verification. On both cases experiment have specified particular sample on different age segment and also shows 76% and 72% overall accuracy rate respectively. The overall accuracy results show that the system can assist the human experts with a great extent.

In chapter 7 automatic authentication of paper money has been targeted. Indian bank notes are taken as reference to show how a system can be developed for discriminating fake notes from genuine ones. Image processing and pattern recognition techniques are used to design the overall approach. Our feature extraction process considers four different aspects of the banknote: i) printer identification ii) Ink property iii) Thread, and (iv) the Art work used in designing the note. The feature space is analyzed and visualized by using clustering technique. The decision making process is built using two different classifiers: i) Artificial Neural Networks (ANNs) ii) Support Vector Machines. Under flow of experiments firstly concerted on thread based features. The two thread based features are therefore used upfront that are binary in nature and discriminate notes based on either the presence or absence of two features. This reduces the processing time too which we have analysed next. Then separately three different areas achieved 99%, 98.5%, 100% by SVM classifier and 99.3%, 98.5%, 99.8% by NN classifier on artwork, ink-based and printing base features respectively. Furthermore, a Linear Discriminate Analysis is used to measure the performance of each feature.

## 8.3 Future Scope of Work

Security art work base authentication provides primary verification of authenticity because this fine line design uses maximum number of security documents. The proposed method which based on principles of image processing and pattern recognition is essentially needed for processing huge number of documents. The future extension of the microprint line authentication method would conduct experiment on a larger dataset to test the generality and scalability of the proposed method. Analysis of error cases has not been reported here. Finding the reasons for occurring errors and design of a more sophisticated approach to minimize these errors are considered as the next part of the study presented here. Finally, experiments with other kinds of security documents like legal deeds, lottery tickets, tickets for watching different outdoor games, certificates, mark sheets, postal stamps, etc. are to be conducted to establish the well acceptance of the method for authenticating security documents. The IHT study can be also extended in several directions. At first, improvement of inverse halftone images is definitely an area where more research is needed. It is shown here that a two-stage neural net performs better than a single stage architecture. This can be further extended to investigate whether a series of neural net could produce better quality inverse halftone. As the current work is motivated by the practical need of generating original gray tone image in situation where the original image is not available or the printed halftone is in question of copyrighting issue, performance evaluation on a much larger dataset is required to bring out the potential of the present approach in practical scenario. Once established its efficiency on a larger dataset the present method could of good assistance for the forensic scientists, printing engineers and even for restoration of many historical documents. In chapter 4, we are mainly concerned about currency note authentication by printing technique verification. The features used in this experiment are quickly computable and therefore, the proposed method provides a quick (and possibly low cost too) solution to the problem. Three different classifiers have been used in this study. Integration of them can also be done in future to build a more robust classification scheme. However, testing of this method using a large dataset is required to establish it as a standard practice. The method, indeed, is not only restricted to currency notes only.

In the chapter 5 reports an interesting experiment in the context of machine aided authentication of security paper documents. This study investigated the role of paper pulps in order to

authenticate a paper. Here we used few number of real dataset i.e. Indian bank notes. The future of this research would consider a separate dataset to test the generality of the approach. The proposed approach could be a generic framework or a part for authenticating any security paper documents where the paper itself carries security aspects. For this purpose, certificates, legal deeds, mark sheets, etc. is to be considered next.

Ink age determination (chapter 6) of printed documents started back in 1920. Iron gallotannate inks properties have determinate the age on that period. From that time, ink dating experiment have great demand for criminal investigations and civil litigations. Capabilities of this experimental knowledge are now widespread throughout the world. In the United State, ink dating on question documents has become a routine question. Early ink dating research is available in the forensic examination of ink and paper. That types of experiments were destructive in nature. That examination was not reasonably cooperative for those documents which require in future. So, our proposed model tries to integrate this scarcity. However, a lot more experiments are to be done to establish this practice as an acceptable one. In fact, the findings of this study open up new research avenues in this domain. Rigorous experiment on a larger dataset and then establishing the statistical significance of the results are to be conducted next. It would be exciting to attempt to compare with other optimization model or pattern recognition techniques for printer inks and then estimate the aging parameters from samples. This model can later be employed to predict the age of a printed document. Similar attempt is reported from chemical viewpoint. And also there might be some technique through destructive and time consuming method which is presently used such type of problem. Our study was lookout to overcome such problem. This scheme has relevancy and will get real challenge in this research area. We believe that future research work would proceed elaborately towards this direction. Extension of the present study for handwriting manuscripts would also be another motivating area of study.

In chapter 7 automatic method for authentication of currency notes is explored. This research is particularly important when the problem of fake is considered as a serious problem in many countries. The present experiment considers Indian bank notes as reference as in the recent times Indian economy has been severely challenged by the presence of fake notes. Moreover, Indian currency notes are nice examples of embedding many heterogeneous security features. The complexity of the overall system is kept optimal so that a low cost hardware realization of the proposed method is feasible. A low cost system is in demand so that a large scale deployment

of such a system becomes possible. For this purpose, we are in touch with a few companies who are interested in prototyping such a system. Some algorithmic optimization may be needed for embedded realization of the present system.

Another immediate extension of this study is to evaluate the method on a different test collection. We are in process of collecting a new set of sample from another laboratory (different from the one from which we received the current data set) of the Department of Forensic Sciences. Exploiting new features and method for authentication is indeed needed to make the system robust against future counterfeiting efforts. In fact, the present study does not consider one important security feature, namely the watermark feature of the currency notes. The reason behind this refers to the strange habits of Indian people scribbling by ink pen over the blank region on the note where watermark is embedded. Such scribbling marks make the use of the watermark feature very sensitive in authenticating bank notes. Our future effort will explore how to get rid of such scribbling marks and use the embedded watermark as one of the security features.

Criminal activity concerning generation of duplicate copies of certain kinds of security documents has been becoming a potential threat to our civil society. Today's quality and complexity of security features are still used as a major prevention. Whereas, complicated designs are still not possible on commonly available reprographic equipment. It would require emphasizing automatic document inspection. This thesis provides such help of an automatic means for verification of authenticity. All experiments were not only focuses currency note but others documents also. All proposed methods consider design of an efficient but low-cost solution to this problem of detecting duplicate documents so that mass scale deployment of such systems can be feasible. Finally, experiments with other kinds of security documents like legal deeds, lottery tickets, tickets for watching games like football, cricket, certificates, mark sheets, postal stamps, etc. are to be conducted with relevant organizational help, to establish the proposed framework as well as an accepted tool for authenticating security documents.

# Publications out of this work

## Ph D Symposium

1 . **Biswajit Halder**, *"Automatic authentication of printed security documents"*, Ph D. Symposium accepted and presented in *Springer*, 10th International Conference on Information Systems Security (ICISS 2014),16-20th december 2014, IDBRT(Established by Reserve Bank of India), Hydrabad, India.

## Journal

2 . Ankush Roy, **Biswaiit Halder**, Utpal Garain and David Doermann. *"Automatic Authentication of Banknotes"*, in Proc. *Springer*, International journal of document analysis and recognition (IJDAR),Vol. 18, Issue 2, DOI: 10.1007/s10032-015-0246-y, Print ISSN 1433-2833, Online ISSN 1433-2825, 2015. (SCI @Thomson router impact factor 0.857 (2013/2014), 5 years impact factor 1.01 Research Gate)

3 . **Biswajit Halder** and Abhoy C. Mondal, *"Modelling of Ink-Colour degradation on old printed documents"*, in Proc. *InderScience*, International Journal of Computational Vision and Robotics (IJCVR), 2015, ISSN Print: 1752-9131, ISSN Online: 1752-914X, IJCVR-83454 (Accepted now in Press).

## Conference

4 . U. Garain and **B. Halder**, *"On Automatic Authenticity Verification of Printed Security Documents"*, in Proc. IEEE, Indian Conference on Computer Vision, Graphics and Image processing(ICVGIP),pp. 706-713, Bhubaneswar, India, 2008.

5 . U. Garain and **B. Halder**, *"A Framework for Machine Authentication of Security Documents"* in Proc. of 19th All India Forensic Science Conf. (AIFSC 2009),pp. 425 - 433, Gandhinagar, Gujrat, India, 18 - 20th January 2009.

6 . U. Garain and **B. Halder**, *"Machine Authentication of Security Documents"*, in Proc. Springer, Int. Conf. on Document Analysis and Recognition (ICDAR),pp. 718-722, Bacelona, Spain, 2009.

7 . **B. Halder** and Abhoy Ch. Mondal, *"Color Based Age Determination on Printed Document"*, in Proc. of 1st National Conference on computing and systems (NACCS '10), pp. 42 - 46, Burdwan, W. Bengal, India, 29th January, 2010.

8 . A. Roy, **B. Halder** and U. Garain, *"Authentication of Currency Notes through Printing Technique Verification"*, in Proc. of ACM, Indian Conference on Computer Vision, Graphics and Image processing (ICVGIP),pp. 383-390, Chennai, India, 2010.

9 . **B. Halder** and Abhoy C. Mondal *"Age based Color Level Variation of Ancient Documents"*, in Proc. of 1st International Conference on computing and systems (ICCS '10), pp. 75 -79, Burdwan, W. Bengal, India, 19 - 20th November, 2010.

10 . **B. Halder** and U. Garain, *"Color Feature Based Approach for Determining Ink Age in Printed Documents"*, in Proc. IEEE, Int. Conf. on Patter Recognition (ICPR), Istanbul, Turkey, pp. 3212-3215, 2010.

11 . **B. Halder**, Abhoy C. Mondal and Rajkumer Darbar *"Currency note authentication checking by fluorescent paper pulp"*, in Proc. of 2nd National Conference on computing and systems (NACCS '12), pp. 233 - 235, Burdwan, W. Bengal, India, 15 - 16th march 2012.

12 . **B. Halder**, R. Darbar and U. Garain, *"Investigating the role of fluorescent paper pulp for detecting counterfeit Banknotes"*, in Springer, 5th IAPR Int. Workshop on Computational Forensics (IWCF), pp.13-23, ICPR, Tsukuba, Japan, 2012.

13 . **B. Halder**, U. Garain, R. Darbar, and A. C. Mondal, *"Inverse of Low Resolution Line Halftone Images for Document Inspection"*, in Proc. 6th IAPR Int. Workshop on Computational Forensics (IWCF), Springer (in press), Sweden, 2014.

14 . **B. Halder**, R. Darbar, U. Garain, A. C. Mondal, *Analysis of Fluorescent Paper Pulps for Detecting Counterfeit Indian Paper Money*, in Proc. Springer, 10th Int. Conf. Information Systems Security (ICISS), pp. 411-424, Hyderabad, India, 16 - 20th December 2014.

# References


[1] O. Hilton, "Scientific examination of questioned documents," *Elsevier Science Publishing Co., New York, US*, 1982.

[2] Conway and V. P. James, "A brief history of the american society of questioned document examiners," *Journal of the American Society of Questioned Document Examiners.*, vol. 1(1), pp. 2–3, 1998.

[3] J. Peterson and M. Hickman, "Forensic science practice in the united states," *The Global Practice of Forensic Science, John Wiley & Sons Ltd, DOI: 10.1002/9781118724248. ch28*, 2014.

[4] C. Lampert, L. Mei, and T. Breuel, "Printing technique classification for document counterfeit detection," *Proc. of Int. Conf. on Computational Intelligence and Security*, pp. 639–644, 2006.

[5] C. CorelDraw, "Security design," *tool of CorelDRAW*, 2011.

[6] J. Levinson, "Questioned documents: A lawyers handbook," *Academic Press Inc., San Diego, US, ISBN-13: 978-0124454903*, 2001.

[7] "Procedure manuals," *Directorate of Forensic Science, Ministry of Home Affairs, Govt. of India*, 2010.

[8] R. Bertrand, P. Gomez-Kromer, R. Terrades, P. Franco, and J. Ogier, "A system based on intrinsic features for fraudulent document detection," *IEEE, 12th International Conference on Document Analysis and Recognition, Washington, DC, US*, 2013.

[9] R. L. van Renesse, "Hidden and scambled images - a review," *In Proc. of SPIE Optical Security and Counterfeit Deterrence Techniques IV, San Jose, CA, US*, vol. 4677, pp. 333–348, 2002.

[10] R. L. V. Renesse, "Protection of high security documents - developments in holography to secure the future market and serve the public," *In Proceedings of Holo-Pack, Holo-Print, Vienna, Austria*, 2006.

[11] K. Koppenhaver, "Forensic document examination: Principles and practice," *Humana Press, NJ, US*, 2007.

[12] "Document security," *Edited - Elite Forms, Inc., New Jersey, US*, January 16, 2012.

[13] C. Felsenfeld and G. Bilali, "Check clearing for the 21st century act - a wrong turn in the road to improvement of the u.s. payments system," *FLASH: The Fordham Law Archive of Scholarship and History*, vol. 85 Neb. L. Rev. 52, 2006.

[14] RBI, "Mechanised cheque processing using micr technology-procedural guidelines (abridged)," *Dept. of Information Technology, central Office, Reserved bank of India, Mumbai, India.http://www.rbidocs.rbi.org.in/rdocs/Publications/DOCs/33140.doc.*

[15] K. Bender, "Moneymakers: the secret world of banknote printing," *Wiley-VCH Publishers, Weinheim (Germany)*, 2006.

[16] R. L. V. Renesse, "Paper based document security - a review," *In: IEEE European Conference on Security and Detection, Commonwealth Institute, London, UK*, pp. 75 – 80, 28 - 30 April, 1997.

[17] S. Huang and J. K. Wu, "Optical watermarking for printed document authentication." *IEEE Trans Inf Forensic Security*, vol. 2(2), pp. 164–173, 2007.

[18] T. Chia and M. Levene, "Detection of counterfeit u.s. paper money using intrinsic fluoresecence lifetime." *Optics Express*, vol. 17(24), 2009.

[19] Cantu and A. Antonio, "A sketch of analytical methods for document dating, part ii,thedynamic approach: Determining age dependent analytical profiles," *International Journal of Forensic Document Examiners, Dating Documents Part II.*, pp. 53–17, 1996.

[20] V. N. Aginsky, "Dating and characterizing writing, stamp pad and jet printer inks by gas chromatography/mass spectrometry," *International Journal of Forensic Document Examiners*, vol. 2(2), 1996.

[21] Rendell and W. Kenneth, "Forging history: The detection of fake letters and documents," *University of Oklahoma Press, Norman, OK, US*, p. 112, 1994.

[22] U. Garain, S. K. Parui, T. Paquet, and H. Laurent, "Machine dating of handwritten manuscripts," *Proceedings of the 9th Int. Conf. on Document Analysis and Recognition (ICDAR), at Curitiba, Brazil*, pp. 759–763, September, 2007.

[23] C. Weyermann, "Mass spectrometric investigation of the aging process of ballpoint ink for the examination of questioned documents," *PhD Thesis, Justus-Liebig-University Giessen*, 2005.

[24] L. Stewart, "Ink age determination by volatile component comparison - a preliminary study," *Journal of Forensic Science,*, vol. 30(2), 1985.

[25] A. Vila, N. Ferrer, J. Mantecon, D. Breton, and J. Garcia, "Development of a fast and non-destructive procedure for characterizing and distinguishing original and fake euro notes," *Analytica Chimica Acta, DOI 10.1016/j.aca.2005.11.084*, vol. 559(2), pp. 257–263, 2006.

[26] "Seeing through counterfeit tricks," *Financial pages of La Vanguardia newspaper, Sunday 24th October 2004.*

[27] "Guide for the development of forensic document examination of forensic capacity," *United Nation Publication, ISBN 978-92-1-130299-8.*, July 2010.

[28] I. Amidror, "A new print-based security strategy for the protection of valuable documents and products using moire intensity," *In Proc. of SPIE Optical Security and Counterfeit Deterrence Techniques, San Jose, CA, USA*, vol. 4(4677), pp. 89–100, 2002.

[29] J. V. Beusekom, F. Shafait, and T. Breuel, "Text-line examination for document forgery detection," *In Springer-Verlag, International Journal on Document Analysis and Recognition (IJDAR), DOI 10.1007/s10032-011-0181-5*, pp. 1–19, 2012.

[30] P. J. Smith, P. O'Doherty, C. Luna, and S. McCarthy, "Commercial anticounterfeit products using machine vision." *In Proc. of SPIE Optical Security and Counterfeit Deterrence Techniques,San Jose, CA, US*, vol. 5310, pp. 237–243, 2004.

[31] N. A. Hampp, M. Neebe, T. Juchem, M. Wolperdinger, M. Geiger, and A. Schmuck., "Multifunctional optical security features based on bacteriorhodopsin," *In Proc. of SPIE Optical Security and Counterfeit Deterrence Techniques.*, vol. 5310, pp. 117–124, 2004.

[32] R. L. van Renesse, "Ordering the order, a survey of optical document security features." *In Proc. of SPIE Conference on Practical Holography IX, San Jose, CA,US*, pp. 268–275, 1995.

[33] F. Garcia-Lamont, J. Cervantes, and Lopez, "Recognition of mexican banknotes via color and texture features." *Expert Systems with Applications*, vol. 39(10), pp. 9651–9660, 2012.

[34] B. Singh, P. Badoni, and K. Verma, "Computer vision based currency classification system," *Int. J. Comput. Appl.*, vol. 16(4), pp. 34–38, 2011.

[35] A. Mikkilineni, P. Chiang, G. Ali, G. Chiu, J. Allebach, and E. Delp, "Printer identification based on texture features," *Proc. of the Int. Conf. on Digital Printing Technologies, Salt Lake City, UT*, pp. 306–311, 2004.

[36] A. Mikkilineni, P. Chiang, and E. Al., "Printer identification based on graylevel co-occurrence features for security and forensic applications," *Proc. of the SPIE Int. Conf. on Security, Steganography and Watermarking of Multimedia Contents VII, San Jose, CA, USA*, pp. 430–440, February, 2005.

[37] N. Khanna, A. K. Mikkilineni, G. T. C. Chiu, J. P. Allebach, and E. J. Delp, "Survey of scanner and printer forensics at purdue university." *Springer LNCS, 2nd Int. Workshop on Computational Forensics (IWCF), Washington, DC, US*, vol. 5158, pp. 22–34, 2008.

[38] A. Mikkilineni, N. Khanna, and E. Delp, "Forensic printer detection using intrinsic signatures," *Proceedings of SPIE-IS and T Electronic Imaging - MediaWatermarking, Security, and Forensics III, San Francisco, CA, US*, 24 - 26 Jan, 2011.

[39] H. Dasari and B. Chakravarthy, "Identification of non-black inks using hsv colour space," *In Proc. of Int. Conf. On Docu. Anal. Recog. (ICDAR), Brazil,*, pp. 486–490, 2007.

[40] G. N. Ali, P. Chiang, A. K. Mikkilineni, J. P. Allebach, G. T. Chiu, and E. J. Delp, "Intrinsic and extrinsic signatures for information hiding and secure printing with electrophotographic devices," *Int. Conf. on Digital Printing Technologies, New Orleans, LA, USA*, pp. 511–515, 28 Sept - 3 Oct, 2003.

[41] M. Schreyer, "Intelligent printing technique recognition and photocopy detection for forensic document examination," *In Proc. of Informatiktage.*, pp. 39–42, 2009.

[42] C. Schulze, M. Schreyer, A. Stahl, and T. Breuel, "Evaluation of gray level features for printing technique classification in high throughput document management systems," *Springer LNCS, 2nd Int. Workshop on Computational Forensics (IWCF), Washington, DC, USA,*, vol. 5158, pp. 35–46, August, 2008.

[43] C. Schulze, M. Schreyer, A. Stahl, and T. M. Breuel, "Using dct features for printing technique and copy detection," *In Proc. of the 5th Int. Conf. on Digital Forensics (IWCF), Orlando, FL, USA*, pp. 95–106, January, 2009.

[44] E. Kee and H. Farid, "Printer profiling for forensics and ballistics," *Proceedings of the 10th ACM Workshop on Multimedia and Security, Oxford, UK*, pp. 3 – 10, 2008.

[45] C. K. Li and S. C. Leung, "The identification of colour photocopiers: A case study," *Journal of the American Society of Questioned Document Examiners*, vol. 1(1), pp. 8–11, 1998.

[46] C. Li, W. Chan, Y. Cheng, and S. Leung, "The differentiation of color laser printers," *Journal of the American Society of Questioned Document Examiners*, vol. 7(2), pp. 105–109, 2004.

[47] J. Tweedy, "Class characteristics of counterfeit protection system codes of color laser copiers," *Journal of the American Society of Questioned Document Examiners*, vol. 4(2), pp. 53–66, 2001.

[48] L. Cui, "Document inspection forged by photocopying." *J. Chinese Peoples Public Secur. Univ. (Science and Technology).*, vol. 3, pp. 22 – 24, 2008.

[49] Y. Wu, X. Kong, and Y. Guo, "Printer forensics based on page documents geometric distortion," *Proceedings of the 16th IEEE International Conference on Image Processing (ICIP), Cairo, Egypt*, pp. 2909 – 2912, 7-12 Nov, 2009.

[50] O. Bulan, J. Mao, and G. Sharma, "Geometric distortion signatures for printer identification," *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1401–1404, 2009.

[51] J. V. Beusekom, F. Shafait, and T. Breuel, "Document inspection using text-line alignment," *ACM International Conference Proceeding Series, Boston, MA, USA*, pp. 263 – 270, 9 - 11 Jun, 2010.

[52] L. Schomaker, M. Bulacu, and K. Franke, "Automatic writer identification using fragmented connected-component contours." *In Proc. of the 9th Int. Workshop on Frontiers in Handwriting Recognition, Tokyo, Japan*, pp. 185–190, October, 2004.

[53] L. Schomaker and M. Bulacu, "Automatic writer identification using connected component contours and edge-based features of uppercase western script." *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 26(6), pp. 787–798, 2004.

[54] G. R. Ball, R. Stittmeyer, and S. N. Srihari, "Writer verification in historical documents," *In Proc. of SPIE Document Recognition and Retrieval*, vol. 7543, pp. 1–8, 2010.

[55] S. Chen and S. Srihari, "A new off-line signature verification method based on graph," *In Proc. of the 18th Int. Conf. on Pattern Recognition (ICPR 06), Hong Kong, China*, pp. 869–872, August, 2006.

[56] D. Pu, G. R. Ball, and S. N. Srihari, "A machine learning approach to offline signature verification using bayesian inference." *Springer, LNCS, 3rd Int. Workshop on Computational Forensics, Hague, Netherlands*, vol. 5718, pp. 125–136, August, 2009.

[57] J. Oliver and J. Chen, "Use of signature analysis to discriminate digital printing technologies," *Proc. of the Int. Conf. on Digital Printing Technologies. San Diego, California*, pp. 218–222, 2002.

[58] A. Burger, "The devils workshop: a memoir of the nazi counterfeiting operation," *Frontline Books, Barnsley*, 2009.

[59] E. Gotaas, "Sensor for verification of genuineness of security paper," *US Patent 5,122,754, June 16, 1992.*

[60] S. Harbaugh, "Capacitive verification device for a security thread embedded within currency paper," *US Patent 5,417,316, May 23, 1995.*

[61] E. Slepyan, A. Kugel, and J. Eisenberg, "Currency verification," *US Patent, No. 6,766,045, July 20, 2004.*

[62] "High level rbi group suggests steps to check menace of fake notes," *Reserve Bank of India, Press release: 2009-2010/232*, 11 August, 2009.

[63] R. L. van Renesse, "Paper based document security - a review." *IEEE European Conference on Security and Detection*, pp. 75 – 80, April, 1997.

[64] J. Takalo, J. Sampo, M. Rantala, S. Siltanen, and M. Lassas., "Using the fibre structure of paper to determine authenticity of the documents: Analysis of transmitted light images of stamps and banknotes," *International Forensic Science, DOI: 10.1016/j.forsciint.2014.09.002*, pp. 252 – 258, 2014.

[65] C. Neumann, R. Ramotowski, and T. Genessay, "Forensic examination of ink by highperformance thin layer chromatography," *J Chromatography the United States Secret Service Digital Ink Library*, vol. 1218(19), pp. 2793 – 2811, 2011.

[66] H. Bugler, H. Buchner, and A. Dallmayer, "Age determination of ballpoint pen ink by thermal desorption and gas chromatography-mass spectrometry," *In Journal of Forensic Sciences.*, vol. 53 (4), 2008.

[67] K. Yoshida, M. Kamruzzaman, F. Jewel, and R. Sajal, "Design and implementation of a machine vision based but low cost stand-alone system for real time counterfeit bangladeshi bank notes detection," *Proc. 10th Int. Conf. on Computer and Information Technology (ICCIT), Dhaka*, pp. 1–5, December, 2007.

[68] A. Guedes, M. Algarra, A. Prieto, B. Valentim, V. Hortelano, S. Neto, R. Algarra, and F. Noronha, "Raman microspectroscopy of genuine and fake euro banknotes." *Int J Rapid Commun, Spectrosc Letter*, pp. 569 – 576, 2013.

[69] C. Seropyan, "Means of preventing counterfeiting bank notes," *US Patent, No. 17,473*, January 2, 1857.

[70] F. Takeda and S. Omatu, "Recognition system of us dollars using a neural network with random masks," *in Proc. of the Int. Joint Conf. on Neural Networks, Nagoya, Japan*, vol. 2, 1993.

[71] A. Frosini, M. Gori, and P. Priami, "A neural network-based model for paper currency recognition and verification," *IEEE Transactions on Neural Networks*, vol. 7(6), pp. 1482–1490, 1996.

[72] M. Aoba, T. Kikuchi, and Y. Takefuji, "Euro banknote recognition system using a three layered perceptron and rbf networks," *IPSJ Trans. on Mathematical Modelling and Its Applications*, vol. 44, pp. 99–109, 2003.

[73] K. Debnath, S. Ahmed, and M. Shahjahan, "A paper currency recognition system using negatively correlated neural network ensemble," *Journal of Multimedia*, vol. 5(6), pp. 560–567, 2010.

[74] H. Hassanpour and P. Farahabadi, "Using hidden markov models for paper currency recognition," *Expert Systems with Applications, Elsevier*, vol. 36(6), pp. 10 105–10 111, 2009.

[75] M. Massimo, "Device for validating banknotes," *EPO Patent, No. EP 0537513 (A1), April 21*, 1993.

[76] B.T.Graves, W.J.Jones, D.U.Mennie, and F.M.Sculits, "Method and apparatus for authenticating and discriminating currency," *US Patent, No. 5,960,103, September 28,1999.*

[77] "Features and utility of paradigm exc 6700-i," *E-Brochure, Paradigm Cash Systems Pvt. Ltd.*, 2009.

[78] "Detection and impounding of counterfeit notes," *Reserve Bank of India, Master Circular*, July 1, 2011.

[79] "Fortuna users guide," *Part-I, Barco Graphics, Gent, Belgium.*

[80] R. D. Warner, R. Adams, and M. Believe, "Introduction to security printing," *Graphic Arts Center Publishing Company, ISBN 0883623757*, 2005.

[81] M. Dorigo and G. Crgo, "The ant colony optimization meta-heuristic," *Editors: New Ideas in Optimization, McGraw- Hill*, pp. 11–32, 1999.

[82] R. O. Duda, P. E. Hart, and D. G. Stork, "Pattern classification (2nd edition)," *Wiley-Interscience, Hoboken, ISBN 0-471-05669-3*, 2000.

[83] J. J. Rodríguez, C. J. Alonso, and J. A. Maestro, "Support vector machines of interval-based features for time series classification," *Knowledge Based System*, vol. 18(4-5), pp. 171–178, 2005.

[84] R. P. Lippmann, "An introduction to computing with neural nets," *IEEE ASSP Magazine*, vol. 4, pp. 4–22, April, 1987.

[85] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," *Psychological Review*, vol. 65(6), pp. 386–408, Nov 1958.

[86] A. A. Desai, "Gujarati handwritten numeral optical character reorganization through neural network," *Pattern Recognition*, vol. 43, Issue 7, pp. 2582–2589, July 2010.

[87] M. J. Aitkenhead and A. J. S. McDonald, "A neural network face recognition system," *Engineering Applications of Artificial Intelligence*, vol. 16, Issue 3, pp. 167–176, April 2003.

[88] M. Amini, R. Jalili, and H. R. Shahriari, "Rt-unnid: A practical solution to real-time network-based intrusion detection using unsupervised neural networks," *Computers & Security*, vol. 25, Issue 6, pp. 459–468, September 2006.

[89] G. Dede and M. H. Sazli, "Speech recognition with artificial neural networks," *Digital Signal Processing*, vol. 20, Issue 3, pp. 763–768, May 2010.

[90] D. Pomerleau, "Alvinn: An autonomous land vehicle in a neural network," *Technical Report CMU-CS-89-107, Carnegie Mellon Univ.*, 1989.

[91] T. Wang, J. S. Huang, and Q. S. Gu, "Photometric redshifts of galaxies from sdss and 2mass," *Research in Astronomy and Astrophysics*, vol. 9(4), pp. 390–400, 2009.

[92] R. Duda and P. Hart, "Pattern classiŕcation and scene analysis." *John Wiley & Sons, New York, US*, 1973.

[93] K. Fukunaga, "Introduction to statistical pattern recognition," *Academic press, Boston, US, 2 edition*, 1990.

[94] P. Jain, H. Hermansky, and B. Kingsbury, "Distributed speech recognition us- ing noise-robust mfcc and traps estimated manner features," *In Proc. In- ternational Conference on Spoken Language Processing, Denver, Colorado, USA*, September 2002.

[95] "A glossary of security features of printed document," *Security and Intelligent Business Unit, U. S. Govt. Printing Office.*

[96] "Fake stamp paper scam," *Reported by Hindustan time, India*, 28th June 2007.

[97] C. L. Owens, "Statement before the u.s. senate committee on banking housing, and urban affairs subcommittee on financial services," *Federal Bureau of Investigation, Financial Institution Fraud, Chief Financial Crimes Section, Washington, D. C.*, September 16, 1997.

[98] R. Ulichney, "Digital halftoning," *MIT Press, Cambridge, MA*, 1987.

[99] D. E. Knuth, "Digital halftones by dot diffusion," *ACM Transactions on Graphics (TOG)*, vol. 6, Issue. 4, pp. 245–273, 1987.

[100] D. L. Lau and G. R. Arce, "Modern digital halftoning," *New York: Marcel Dekker*, 2001.

[101] "Counterfeit banknotes," *Report of the parliamentary office of science and technology, UK, www.parliament.uk/briefing-papers/POST-PN-77.pdf*, 1996.

[102] W. B. Huang, A. W. Y. Su, and Y. H. Kuo, "Neural network based method for image halftoning and inverse halftoning," *in Expert Syst. Appl.*, vol. 34, No. 4, pp. 2491 –2501, 2008.

[103] P. C. Chang and C. S. Yu, "Neural net classification and lms reconstruction to halftone images," *in Proc. SPIE Visual Communications and Image Processing*, vol. 3309, pp. 592–602, 1998.

[104] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 9(1), pp. 62–66, 1979.

[105] A. Hore and D. Ziou, "Image quality metrics: Psnr vs. ssim," *in Proc. IEEE, 20th Int. Conf. Pattern Recognit (ICPR 2010), Istanbul, Turkey*, pp. 2366–2369, 23-26 August 2010.

[106] A. Foi, V. Katkovnik, K. Egiazarian, and J. Astola, "Inverse halftoning based on the anisotropic lpa-ici deconvolution," *in Proc. Int. TICSP workshop spectral meth. multirate signal proc.(SMMSP), Vienna*, pp. Pages:49–56, 2004.

[107] R. Neelamani, R. Nowak, and R. Baraniuk, "Winhd: Wavelet-based inverse halftoning via deconvolution," *Rejecta Mathematica*, vol. 1, pp. 84 – 103, july 2009.

[108] R. Stevenson, "Inverse halftoning via map estimation," *in IEEE Trans. Image Processing*, vol. 6, pp. 574 –583, 1997.

[109] M. Mese and P. P. Vaidyanathan, "Look-up table (lut) method for inverse halftoning," *in IEEE Trans. on Image Processing*, vol. 10(10), pp. 1566–1578, 2001.

[110] E. Peli, "Contrast in complex images," *J. Opt. Soc. of Am.*, vol. 7(10), pp. 2032–40, 1990.

[111] R. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Systems, Man and Cybernetics (SMC)*, vol. 3(6), pp. 610–621, 1973.

[112] A. Colorni, M. Dorigo, and V. Maniezzo, "Distributed optimization by antcolonies," *1st European Conf.on Artificial Life (ECAL), Cambridge, MIT Press, Mass, USA*, pp. 134–142, 1991.

[113] E. Speckin, "Ink dating examination," *Journal of National Association of Document Examiners, Inc., Spring*, vol. 23(1), 2000.

[114] V. N. Aginsky, "A microspectrophotometric method for dating ballpoint inks - a feasibility study," *Journal of Forensic Sciences*, vol. 40(3), pp. 475–478, 1995.

[115] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of psnr in image/video quality assessment," *Electronics Letters*, vol. 44(13): 800, doi:10.1049/el:20080522, 2008.

[116] Times, "Life: Dead & alive," *TIME Press, US*, 1936.

[117] A. French, "The very first issues of 19 famous magazines," *Mental Floss*, 12 August 2013.

[118] "Google makes life magazine photo archives available to the public," *Ewen MacAskill in Washington*, November 18, 2008.

[119] "Counterfeit currency in canada," *Publication of Royal Canadian Mounted Police, Canada*, December, 2007.

[120] R. Kaushal, "Fake money circulation boosts black economy," *Reported - India Today*, August 5, 2009.

[121] R. Gonzalez and R. Woods, "Digital image processing," *Prentice Hall, ISBN number 9780131687288.*, 2008.

[122] R. Duda, P. Hart, and D. Strok, "Pattern classification," *John Wiley & Sons Inc., ISBN: 978-0-471-05669-0*, vol. 18, No.2, 2001.

[123] J. Kelly and B. Lindblom, "Scientific examination of questioned documents," *CRC Press: Boca Raton, FL, US*, 2006.

[124] R. Kuehni, "Color space and its divisions: Color order from antiquity to the present," *John Wiley and Sons*, 2003.

[125] K. Franke and S. Rose, "Ink-deposition model: The rela- tion of writing and ink deposition processes," *In Proc. of the Ninth International Workshop on Frontiers in Handwriting Recognition (IWFHR)*, pp. 173–178, 2004.

[126] "Visual spectral comparator 5000, (vsc 5000)," *Available at: http:// crimesight.co.za.*

[127] J. Duchene and S. Leclercq, "An optimal transformation for discriminant and principal component analysis," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 10(6), pp. 978 – 983, 1988.