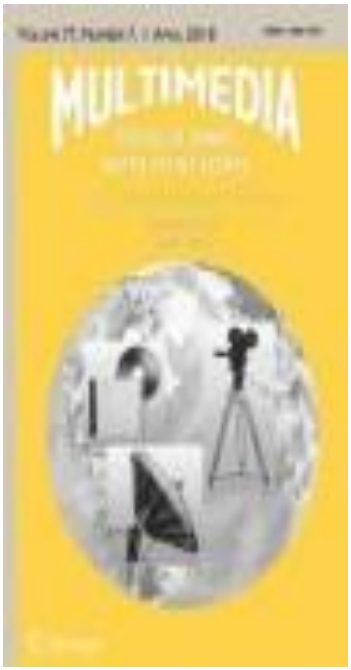

MULTIMEDIA

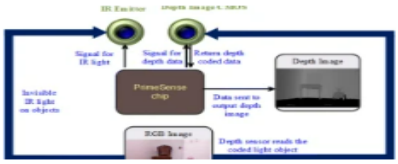
Aims and scope

Multimedia Tools and Applications publishes original research articles on multimedia development and system support tools, and case studies of multimedia applications. Experimental and survey articles are appropriate for the journal. The journal is intended for academics, practitioners, scientists and engineers who are involved in multimedia system research, design and applications. All papers are peer reviewed.

Editors

Shu-Ching Chen, *Florida International University, Miami, Florida, USA*
Tat-Seng Chua, *National University of Singapore*
Ernesto Damiani, *University of Milano, Italy*
Antitza Dantcheva, *INRIA, France*
Chabane Djeraba, *LIFL Laboratory, France*
Schahram Dustdar, *Vienna University of Technology, Wien, Austria*
Abdulmotaleb El Saddik, *University of Ottawa, Canada*
Ben Falchuk, *Vencore Labs, USA*
Marco Furini, *University of Modena and Reggio Emilia, Italy*
Ombretta Gaggi, *University of Padua, Italy*
David C. Gibbon, *AT&T Labs - Research, USA*
Yihong Gong, *NEC, USA*
Sonja Grgic, *University of Zagreb, Croatia*
M. Shamim Hossain, *King Saud University, Saudi Arabia*
Kien A. Hua, *University of Central Florida, USA*
Benoit Huet, *Institut Eurecom, France*
Horace H.S. Ip, *City University of Hong Kong*
Ebroul Izquierdo, *Queen Mary, University of London, United Kingdom*
Ramesh Jain, *University of California, Irvine, USA*
Rongrong Ji, *Xiamen University, China*
Shuqiang Jiang, *Chinese Academy of Sciences, China*
Hari Kalva, *Florida Atlantic University, Boca Raton, Florida, USA*
Shunsuke Kamijo, *University of Tokyo, Japan*
Ahmed Karmouch, *University Ottawa, Canada*
Harald Kosch, *University of Passau, Germany*
Clement Leung, *Hong Kong Baptist University, Hong Kong*
T.D.C. Little, *Boston University, USA*
Huadong Ma, *Beijing University of Posts & Telecommunications, PR of China*
Hong Man, *Stevens Institute of Technology, Hoboken, New Jersey, USA*
Oge Marques, *Florida Atlantic University, Boca Raton, Florida, USA*
Liam M. Mayron, *Florida Institute of Technology, Florida, USA*
Michele Nappi, *University of Salerno, Italy*
Vincent Oria, *New Jersey Institute of Technology, New Jersey, USA*
Charles B. Owen, *Michigan State University, USA*
Dragutin Petkovic, *San Francisco State University, USA*
Giuseppe Polese, *University of Salerno, Italy*
Jorge Posada, *Vicomtech Research Centre, Spain*

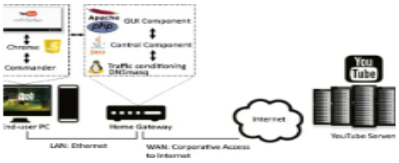
Improved visual SLAM: a novel approach to mapping and localization using visual landmarks in consecutive frames



Kajal Sharma

OriginalPaper | Published: 19 April 2017 | Pages: 7955 - 7976

Analysis of YouTube's traffic adaptation to dynamic environments



Javier Añorga, Saioa Arrizabalaga ... Jaizki Mendizabal

OriginalPaper | Published: 27 April 2017 | Pages: 7977 - 8000

Segmentation of brain MR images using a proper combination of DCS based method with MRF

Ali Ahmadvand, Mohammad Reza Daliri & Sayyed Mohammadreza Zahiri

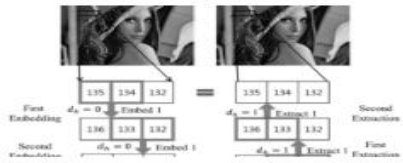
OriginalPaper | Published: 09 May 2017 | Pages: 8001 - 8018

High-quality blind watermarking in halftones using random toggle approach

Yung-Yao Chen & Wei-Sheng Chen

OriginalPaper | Published: 13 April 2017 | Pages: 8019 - 8041

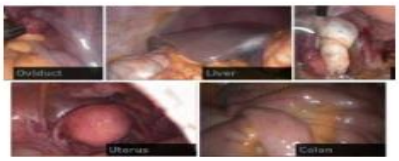
Genuine reversible data hiding technology using compensation for H.264 bitstreams



Hyunjung Kim & Sang-ug Kang

OriginalPaper | Published: 27 April 2017 | Pages: 8043 - 8060

Learning laparoscopic video shot classification for gynecological surgery



Stefan Petscharnig & Klaus Schöffmann

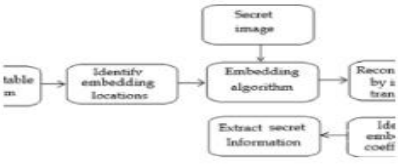
OriginalPaper | Open Access | Published: 22 April 2017 | Pages: 8061 - 8079

Effective client-driven three-level rate adaptation (TLRA) approach for adaptive HTTP streaming

Selvaraj Kesavan & J. Jayakumar

OriginalPaper | Published: 09 May 2017 | Pages: 8081 - 8114

Curvelet transform and cover selection for secure steganography



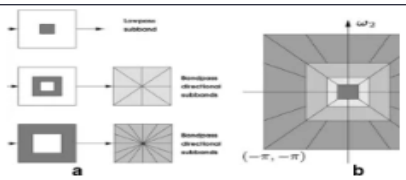
Mansi S. Subhedar & Vijay H. Mankar

OriginalPaper | Published: 30 May 2017 | Pages: 8115 - 8138

[Video shot boundary detection using multiscale geometric analysis of nsct and least squares support vector machine]

Jaydeb Mondal, Malay Kumar Kundu ... Manish Chowdhury

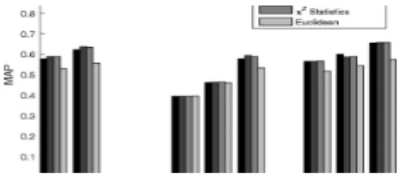
OriginalPaper | Published: 25 April 2017 | Pages: 8139 - 8161



[Improving content-based image retrieval for heterogeneous datasets using histogram-based descriptors]

Carolina Reta, Ismael Solis-Moreno ... Paul Townend

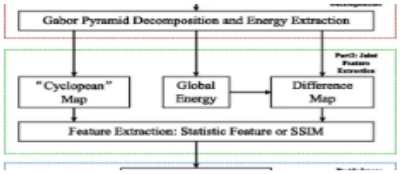
OriginalPaper | Published: 01 May 2017 | Pages: 8163 - 8193



[No-reference stereoscopic 3D image quality assessment via combined model]

Lili Shen, Jinyi Lei & Chunping Hou

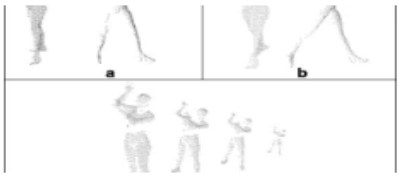
OriginalPaper | Published: 21 April 2017 | Pages: 8195 - 8212



[Action recognition from point cloud patches using discrete orthogonal moments]

Huaining Cheng & Soon M. Chung

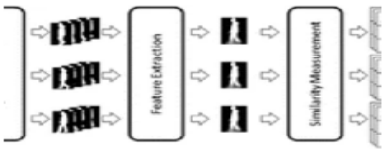
OriginalPaper | Published: 27 April 2017 | Pages: 8213 - 8236



[Clothing-invariant human gait recognition using an adaptive outlier detection method]

A. Ghebleh & M. Ebrahimi Moghaddam

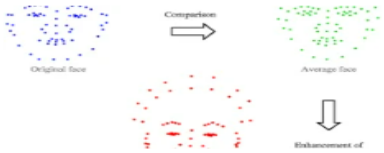
OriginalPaper | Published: 09 May 2017 | Pages: 8237 - 8257



[Caricature generation utilizing the notion of anti-face]

Vlasis Gogousis & Anastasios Tefas

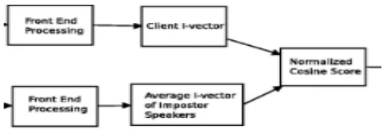
OriginalPaper | Published: 08 May 2017 | Pages: 8259 - 8271



[Client-wise cohort set selection by combining speaker- and phoneme-specific I-vectors for speaker verification]

Waquar Ahmad, Harish Karnick & Rajesh M. Hegde

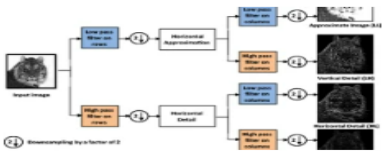
OriginalPaper | Published: 24 May 2017 | Pages: 8273 - 8294



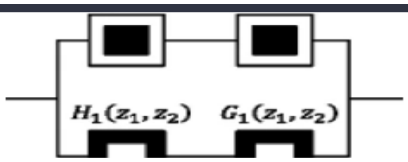
[Maximizing embedding capacity *and* stego quality: curve-fitting in the transform domain]

Tamer Rabie, Ibrahim Kamel & Mohammed Baziyad

OriginalPaper | Published: 04 May 2017 | Pages: 8295 - 8326



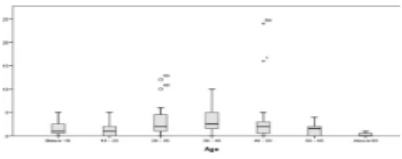
Multi-focus image fusion based on nonsubsampled compactly supported shearlet transform



Chunyu Wei, Bingyin Zhou & Wei Guo

OriginalPaper | Published: 08 May 2017 | Pages: 8327 - 8358

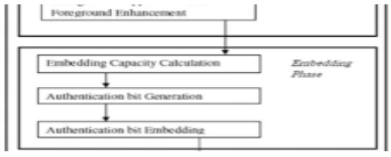
Definition of requirements for accessing multilingual information opinions



Jan Derkacz, Mikołaj Leszczuk ... Kamel Smaïli

OriginalPaper | Open Access | Published: 27 May 2017 | Pages: 8359 - 8374

Self authenticating medical X-ray images for telemedicine applications



Rajitha Bakthula, Shivendra Shivani & Suneeta Agarwal

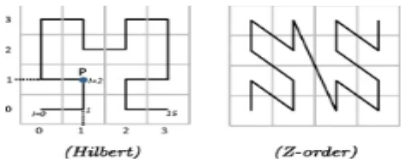
OriginalPaper | Published: 28 April 2017 | Pages: 8375 - 8392

Transreceiving of encrypted medical image – a cognitive approach

Padmapriya Praveenkumar, N. Kerthana Devi ... Rengarajan Amirtharajan

OriginalPaper | Published: 29 April 2017 | Pages: 8393 - 8418

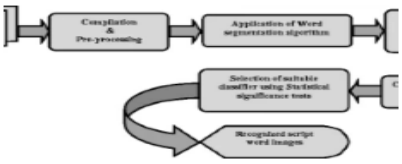
Alternative patterns of the multidimensional Hilbert curve



Patrick Franco, Giap Nguyen ... Jean-Marc Ogier

OriginalPaper | Published: 10 May 2017 | Pages: 8419 - 8440

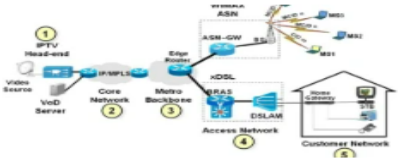
Benchmark databases of handwritten *Bangla-Roman* and *Devanagari-Roman* mixed-script document images



Pawan Kumar Singh, Ram Sarkar ... Mita Nasipuri

OriginalPaper | Published: 18 May 2017 | Pages: 8441 - 8473

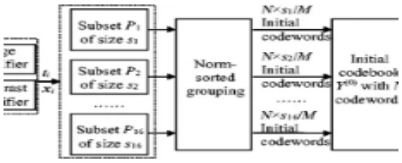
On the synthetic dataset generation for IPTV services based on user behavior



Alireza Abdollahpouri, Reyhan Qavami & Parham Moradi

OriginalPaper | Published: 10 May 2017 | Pages: 8475 - 8493

VQ codebook design using modified K-means algorithm with feature classification and grouping based initialization



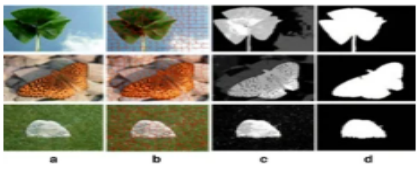
Lang Wang, Zhe-Ming Lu ... Ya-Pei Feng

OriginalPaper | Published: 20 May 2017 | Pages: 8495 - 8510

[Performance enhancement of salient object detection using superpixel based Gaussian mixture model](#)

Navjot Singh, Rinki Arya & R. K. Agrawal

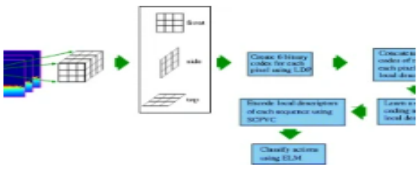
OriginalPaper | Published: 20 May 2017 | Pages: 8511 - 8529



[Local derivative pattern for action recognition in depth images](#)

Xuan Son Nguyen, Thanh Phuong Nguyen ... Ngoc-Son Vu

OriginalPaper | Published: 04 May 2017 | Pages: 8531 - 8549



[Text recognition in scene image and video frame using Color Channel selection](#)

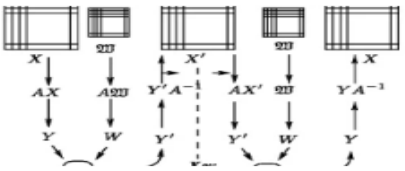
Ayan Kumar Bhunia, Gautam Kumar ... Umapada Pal

OriginalPaper | Published: 05 May 2017 | Pages: 8551 - 8578

[More secure lossless visible watermarking by DCT](#)

Yih-Kai Lin, Cheng-Hsing Yang & Jinn-Tsong Tsai

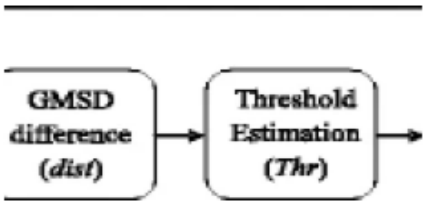
OriginalPaper | Published: 05 May 2017 | Pages: 8579 - 8601



A robust and secure key-frames based video watermarking system using chaotic encryption

Yassine Himeur & Abdelkrim Boukabou

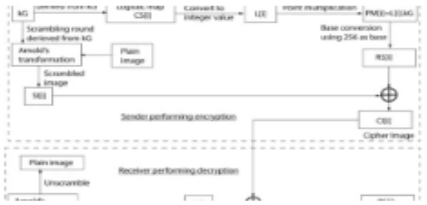
OriginalPaper | Published: 06 May 2017 | Pages: 8603 - 8627



A robust image encryption scheme based on chaotic system and elliptic curve over finite field

Dolendro Singh Laiphrakpam & Manglem Singh Khumanthem

OriginalPaper | Published: 05 May 2017 | Pages: 8629 - 8652



No-reference perceived image quality measurement for multiple distortions

Anupama B. Lamb & Madhuri Khambete

OriginalPaper | Published: 17 May 2017 | Pages: 8653 - 8675

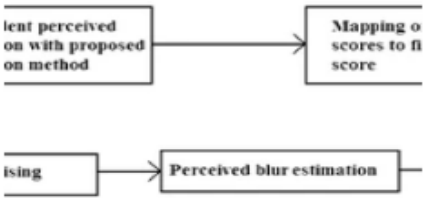
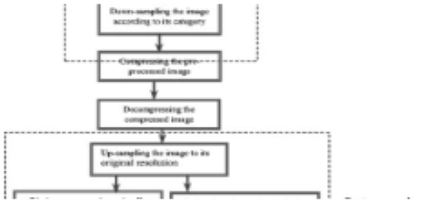


Image compression using JPEG with reduced blocking effects via adaptive down-sampling and self-learning image sparse representation

Sekine Asadi Amiri & Hamid Hassanpour

OriginalPaper | Published: 12 May 2017 | Pages: 8677 - 8693



Multimed Tools Appl (2018) 77:8729–8758
DOI 10.1007/s11042-017-4771-1

Digital forensics of microscopic images for printed source identification

Min-Jen Tsai[1] · **Imam Yuadi**[1,2]

Received: 13 August 2016 / Revised: 24 April 2017 / Accepted: 28 April 2017 /
Published online: 23 May 2017

Abstract When trying to identify a printed forged document, examining digital evidence can prove to be a challenge. In this study, microscopic images are used for printed source identification due to their high magnification properties resulting in detailed texture and structure information. Prior research implemented a scanner as a digitizing technique to resolve very fine printed identification, but this technique provided limited information on the resolution and magnification of the sample. In contrast, the performance of microscopy techniques can retrieve the shape and surface texture of a printed document with differing micro structures among printer sources. To explore the relationship between source printers and images obtained by the microscope, the proposed approach utilizes image processing techniques and data exploration methods to calculate many important statistical features, including: Local Binary Pattern (LBP), Gray Level Co-occurrence Matrix (GLCM), Discrete Wavelet Transform (DWT), Spatial filters, the Wiener filter, the Gabor filter, Haralick, and SFTA features. Among the different set of features, the LBP approach achieves the highest identification rate and is significantly superior to other methods. As a result, the proposed technique using microscopic images achieves a high classification accuracy rate, which shows promising applications for real world digital forensics research.

Keywords Microscopic images · Digital image forensics · Feature filters · Support vector machines (SVM) · Local binary pattern (LBP)

✉ Min-Jen Tsai
mjtsai@cc.nctu.edu.tw

[1] Institute of Information Management, National Chiao Tung University, 1001 Ta-Hsueh Road, Hsin-Chu 300 Taiwan, Republic of China

[2] Department of Information and Library Science, Airlangga University, Jl. Airlangga 4-6, Surabaya, East Java 60286, Indonesia

1 Introduction

Digital forensics is the examination and analysis of digital evidence to prove the occurrence of a crime. Digital forensics implements specific tools and methods to identify, collect, and analyze digital evidence [28]. Recently, digital forensics for printed document source identification has begun to be increasingly important in the investigation and prosecution of many types of crimes. The challenges in the field of forensic investigation still rise to provide appropriate and sufficient security measures and devices [10, 25] in the forensic process to help forensic investigation. However, these devices give the potential effects where the digital documents often contain information about crimes committed, movement of suspects, and hidden messages. Correspondingly, documents in a suspect's possession might possibly reveal clues from digital evidence. Using digital evidence in a legal trial can prove to be challenging. Therefore, it requires accurate techniques to prove authenticity of digital evidence. Digital forensics experts are needed to assist law enforcement to determine whether a suspect is guilty or innocent of a crime, by examining evidence using standard investigation techniques as well as a broad range of digitizing tools such as cameras, scanners, and microscopes.

The development of methods and tools by digital forensics researchers in recent years has led to significant improvements in the forensic sciences. Advanced techniques and materials have been developed to make results more accurate. For instance, different digitizing document devices, such as scanners, cameras, or microscope instruments, can provide detailed image resolution for printed source identification. Digitizing techniques or imaging documentation is the translation of a printed document into binary data. It is the representation of an object, image, or document with a series of numbers that describe a discrete set of points or samples [19]. Digitizing a document plays an important role in the printed source identification. Generally, the process of digitizing documents with or without magnification is widely accepted in research. For example, a scanner is a universal device that optically scans images, printed texts, handwriting, or objects for digital format without magnification [1, 4–6, 12, 23, 30, 31, 37, 40, 45–47, 49]. On the other hand, the microscope is used for micro scale imaging with magnification to capture the details of an object, image, or document to identify speckle, ink and toner printers [3, 7, 33–36, 41, 42, 44, 50].

Most studies of micro scale imaging of printed documents have focused primarily on recognizing the chemical composition of printer toner used or speckle pattern of paper. By contrast, studies of printed source identification by using microscopic images are less based on machine learning classification by the researchers. The aim of this study is to find the best solution in identifying the source printer by expanding feature extraction through the microscopic images. The question is: by using a microscope to analyze a printed document identifying it from different printers, can the identification method be improved? This study seeks to:

a. Develop different microscopic image techniques for text and image documents
b. Identify laser printer sources using microscopic images based on the SVM classification
c. Analyze images by using different feature extractions to attain accurate identification

This paper is organized as follows: Section 2 describes the related works and feature filters. Section 3 presents the proposed approach used in this study. Section 4 details the results with discussion, and Section 5 concludes the paper with areas of possible future investigation.

2 Related works and feature filters

Before describing the proposed method of this research, we will discuss related solutions for laser printer identification. We will also discuss the advantages and limitations of using the scanner as digitizing device. Subsequently, the mechanism of the microscope used to digitize the printed document will be explained. In addition, the adopted features for image characteristic extraction will be systematically examined based on the microscopic images.

2.1 Related works

Several researchers employed scanner systems in their forensics process. Mikkilineni [30] used a scanner with 2400 dpi with 8 bits/pixel to digitize English character "e" for printer source identification. Ryu et al. [37] also used 2400 dpi scanner to investigate the halftone texture of a color document by employing histograms of angles from Hough transform for each CMYK band. Furthermore, Ali et al. [1] applied Gaussian mixture model and binary tree as the classifiers in order to reduce the dimension of the data set from multiple projections to identify the printer. Tsai et al. [46] set up the scanner resolution as low as 300 dpi and 600 dpi for Chinese character source identification, which achieved high accuracy rate using appropriate feature extraction in the machine learning environment. In specific purposes, a scanner can be used in forensic processes for physical object authentication [40], and image based data interfaces revisited [49]. On the other hand, the scanner technique still same limitations, especially, the details of texture information. For this reason, micro scale forensic imaging can provide extra knowledge towards different approaches and subsequent analysis during the investigation, such as microscope, mid-infrared (IR), near-infrared (NIR) and Raman chemical imaging. The imaging systems that have correlation in microscopic images in printed document identification, Polard et al. [35] presented a model-based approach for extracting a signature profile around the outer edge of virtually any text glyph. They used two high-resolution imaging devices i.e., the Dyson Relay CMOS Imaging Device, called DrCID, and a high speed line-scan camera. This signature encodes that part of the glyph boundary which is due to the random fluctuation of the print process. By using the same device, Simske and Adams [42] also analyzed of the single printed character to simultaneously provide fiducial marking, inspection information, authentication and forensics. They implemented custom forensic shape analysis software (FSAS) to identify individual text characters based on segmentation, shape and text print quality assessment during the forensic process. Meanwhile, the forensic researches can be performed through chemical approaches by using microscopes and chemical imaging. Božičević et al. [36] implemented chemical imaging by using micro-Raman spectroscopy to identify the common origin of toner printed counterfeit. In addition, principal component analysis (PCA) and laser ablation inductively coupled with plasma mass spectrometry were demonstrated in forensic analysis for laser printer with toner powder [7, 44] and ink [33].

Applying microscopes as digitizing tools to assist an investigation of printed document enables us to acquire question images in details. By utilizing the inherent non-repeatable randomness existing in a printing process, Zhu et al. [50] used electronic or USB microscope to analyze print signatures from different kinds of paper. Similarly, Sharma et al. [41] also applied the microscope to evaluate fingerprints across different types of paper and conditions. They extracted texture paper speckle patterns from regions of papers using Gabor Transform and Singular Value Decomposition (SVD) to generate fingerprint. To acquire a speckle pattern

of the investigated object, the laser speckles have been used by Buchanan et al. [3] to investigate the intrinsic roughness as the physical properties present on all non-reflective surfaces as a source of physical randomness. By using a line-shaped laser focus, four photo detectors and a scanning mechanism, the paper sample was recognized in their speckle pattern. Osadchy et al. [34] presented that the non-isotropic surfaces were a number of image representations equivalent to, or closely related to the image gradient direction. It indicates the texture and surface dependent representation for illuminating insensitive image comparisons.

In machine learning environment, feature extraction is the most important stage after the document has been digitized into the numeric representation. Mikkilineni et al. [30, 31] applied gray level co-occurrence matrix (GLCM) for each "e" character to form the feature vectors. Several researchers in [6, 12, 45–47] also conducted their experiments by using GLCM along with different approaches and extended features. Tsai et al. [45, 46] implemented GLCM and discrete wavelet transform (DWT) based feature extraction to identify Chinese character, and used feature selection to achieve the optimum classification of printer source identification. In a further study [47], they identified Japanese character with more features, which include GLCM, DWT, Gaussian, LoG, Usharp, Wiener and Gabor features. Ferreira et al. [12] used several features such as GLCM [6] variations, HOG, LBP and others to identify printed document "e" and frame document. Furthermore, Kim and Lee [23] applied each CMY color channel in the discrete Fourier transform (DFT) domain to identify the color laser printer. They applied 15 halftone texture features that were extracted from the preprocessed images before classification. Correspondingly, Bulan et al. [4] also assessed the similarity of a pair of geometric distortion signatures during the printing process using the normalized correlation.

2.2 The advantages and the limitations of using scanners

A scanner is a device that is able to convert texts or printed documents into digital formats, which has now become an important part of the home office with easy-to-use software support over the last few years. There are several advantages to use the scanner for forensics application, such as portability, cost effectiveness, speediness, efficiency, etc. By using this device, virtually all kinds of documents can be captured in minutes with high accuracy. On the other hand, there are several disadvantages to use the scanner in printed source investigation. For example, the magnification limitation of the scanner makes it unable to obtain detailed texture information at the microscopic scale. The image resolution is not sufficient for providing details on the surface features: the shape, the size and the structure for the micro surface texture.

Several prior researches applied scanners to digitize printed documents in printer source investigation, the result of which are tabulated in Table 1. These scanners convert the hard-copy documents into digital data, and help to create the original, the authentic, and the chain of custody of such digital images to the forensic examiner [5]. By using the scanner for the forensics application, the examiner gets the image sample in digital form, which is generally equal to the size of the original object without magnification. Therefore, microscopes with the magnification system can capture the texture and the pattern object, including its size, its shape, and its color composition in micro-scale imaging. With much improved image texture information, this study expects that the substantial knowledge can help to improve the accuracy rate for the printed source identification.

Table 1 Research resources

Literature	Imaging devices	Approach	Research objective	Classification technique	Claimed accuracy rate	Number of printers
[46]	Sc	GLCM, DWT	Printed document (Chinese character)	SVM	98.64%	12
[31]	Sc	GLCM	Printed document "e"	SVM	93.0%	10
[47]	Sc	GLCM, DWT, Gaussian, others	Printed document (Japanese character)	SVM	94.23%	12
[12]	Sc	GLCM variations, HOG, LBP and others	Printed document "e" and frame document)	SVM	98.47%	10
[6]	Sc	DWT, GLCM	Printed document (color image document)	SVM	99.34%	8
[23]	Sc	Discrete Fourier transform (DFT)	Photograph Image	SVM	94.4%	7
[35]	DrCID	Model based signature profile (MBSP), Shape Warp Coding (SWC). Histogram comparison	Printed document (a,s,A,S,9)	*	*	2
[36]	OM + CI	Raman spectroscopy technique Histogram comparison	a powder or liquid toner	*	*	11
[44]	CI	cluster analysis (CA) and principal component analysis (PCA), Laser ablation inductively coupled-plasma mass-spectrometry	Toner powder	*	Correlation 96–99%	201 black toner and 23 color toner
[7]	XRF + LEAF	Principal component analysis (PCA), Laser ablation inductively coupled-plasma mass-spectrometry, Histogram comparison	Printed document, toner powder	K-Means WEKA	*	4
[33]	CI	NIR spectroscopy, principal component analysis (PCA) & Histogram comparison	Inkjet printers	*	*	19
[50]	USB-M	Radius values and shape matching	. (dot), p in four different kinds of paper	*	*	2 printers and 1 photocopier
[41]	USB-M	Gabor binary matrix and Singular Value Decomposition (SVD)	Fingerprints	*	*	*

1. Sc, DrCID, OM, CI, XRF, LEPF, and USB-M denotes imaging devices by using (Sc)anner, Dyson Relay CMOS Imaging Device that called DrCID, (O)ptical (M)icroscopy, (C)hemical (I)maging or spectrometer, XRF is X-Ray Fluorescence, and (L)aser-(E)xcited (A)tomic (F)luorescence, and USB-(M)icroscopy

2. * denotes no exact information is provided for that item

2.3 How the optical microscopes work

Based on the lighting system, there are two types of optical microscope i.e., reflected light illumination and transmitted illumination. To find out how the optical microscope work, Fig. 1 illustrates the process of two types of optical microscope for image formation. Fig. 1(a) shows the transmitted light microscope (TLM) which is a type of optical microscope where the light is transmitted from a source on the opposite side of the specimen. After the light passes through the specimen, the image of the specimen goes through the objective lens to the oculars or subjective lens, where the enlarged image is viewed by human eyes, or captured by the digital camera. This microscope is for use on the cell, blood or liquid material observation. Conversely, the other type of optical microscope is the reflected light microscope (RLM) which is shown in Fig. 1(b). The light path of RLM comes from the light source, then moves into the excitation filter, and finally passes through the objective lens; is then reflected off the surface of the specimen, and returns through the objective, while finally reaching the eyepiece or the camera for observation [11, 16, 29]. Both of them have different function depending on the material specimen to be observed. The most suitable optical microscope to view printed document in this study will be discussed in Sec. 3.2.3. Thereafter, the chosen type of microscope will be used in the experiments for performance analyses.

In brief, the printing process is an integrated process between the unevenness of the photosensitive surface from the printer, the paper surface, the variable granularity of the toner powder, the unstable heat, the pressure of the fuser, the amount of excess toner remaining on the photoreceptor, and many other such factors. The net outcome of all these variabilities forms the unique signature for each printer. Therefore, the features of the microscopic images from printed document should be systematically extracted and analyzed. Since the techniques developed in previous works can not be directly applied, this study investigates those issues and proposes a unified approach in printer source identification.

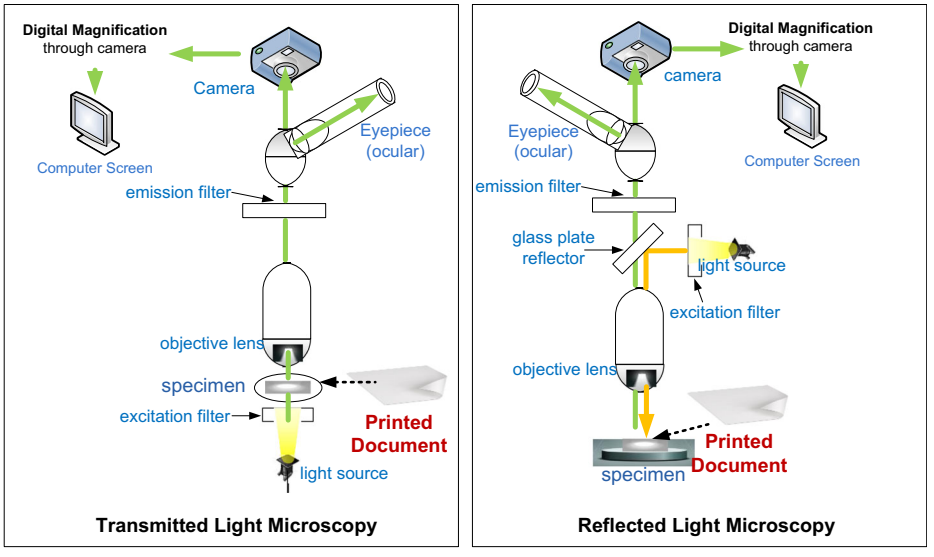


Fig. 1 The optical microscope based on transmitted and reflected light illumination

2.4 The feature filters for the performance statistics

To acquire an initial set of measured data for the derived values or features from each of the microscopic image is known as the feature extraction or filtering. This work used ten different set of filters in order to acquire the most informative values in order to make the best decision. Due to the limit of the space confined here, concise explanation can only be used. Interested readers can refer to [2, 8, 9, 13–15, 27, 32, 39, 48] for more detailed information.

2.4.1 The LBP features

LBP is a gray-scale invariant texture and combination between measuring texture from each neighborhood, and the difference of the average gray level of those pixels based on binary number. It is a feature extractor that has an appropriate and powerful measure of images texture according to the results of empirical studies. The advantage of LBP is that it can obtain pattern labels for microscopic texture images with the histogram of uniform LBP, especially for the printed document texture with different edges and shapes [32]. For unsupervised texture segmentation, LBP operator is used together with the size of the local contrast as simple filter feature, which has high quality performance. The definition of uniform pattern is the extension of the original carriers that can be used to reduce the length of the feature vector and a simple invariant descriptor, as is defined in Appendix Table 12. To implement LBP feature for microscopic images of printed document, a series of experiment process can be conducted. Firstly, we specified the area of character that would be captured by using an optical microscope. For example, the specific area in character "e" is corresponded by the red box in Fig. 2.

Secondly, the printed document in the microscopic image is equally divided into a number of patches that represent the texture from specific printer sources. We examined the microscopic images that are cropped in a certain pixel of patch. Each patch has 43 × 43 pixel size that will be used for a sample image in this research. Thirdly, the sample image is divided into blocks during the LBP feature extraction. Inspired from [27], we implemented LBP (8,1) neighborhood which is LBP pixel within the printed area of character "e". For each pixel in a block, we compared the pixel to each of its 8 neighbors that the direction is from left-top, left-middle, left-bottom, right-top, etc.). When the pixel value is greater than the neighbor's value, write "0". Otherwise, write "1" as the LBP value. Finally, we computed the histogram value from each block, where it is a combination of those pixels which are smaller and the pixels which are greater than the center. There are 2^8 possible combinations with 8 surrounding pixels. The total value of LBP are merged into one histogram when using uniform patterns, and the length of feature vector reduced from 256 to 59 for a single cell, if using R= 1 and P=8. The combined 59 features are further used as feature filters in this study.

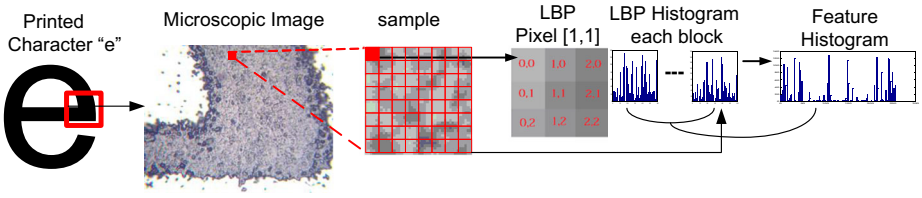


Fig. 2 The LBP feature from printed document "e"

2.4.2 The spatial features

In this study, several spatial feature techniques are adopted for analysis of microscopic images. They are GLCM, Discrete Wavelet Transform (DWT), Laplacian of Gaussian (LoG), Unsharp, Wiener, Gabor, and Haralick filters. These filters are briefly defined in Appendix Table 12 and described here as the following:

1) GLCM features are the estimation of the second order probability density function of the pixels in the image and the features are statistics obtained from the GLCM [30]. There are four directions to generate the data that could be focused on during the generation of the matrix, and these are: 0 degrees (horizontal direction), 45 degrees, 90 degrees (vertical direction), and 135 degrees. The direction and spatial distance from the reference pixel *i* will be defined, such as 1 space at 45 degrees direction locates the adjacent pixel *j*, next to the reference pixel *i* [30, 46]. Consequently, (*i*, *j*) indicates the spatial location of image; *glcm (n, m)* means the number of occurrences of pixels with gray levels *n* and *m* respectively with a separation of (*dr,dc*) pixels. If the GLCM is normalized with respect to *Rglcm*, its entries then represent the probability of the occurrence of pixel pairs, with graylevel *n* and *m* with separation (*dr,dc*). Here we choose $dc = 0$ and $dr = 1$. The formula is defined in the Appendix Table 12. It is generated as a binary image map with all the pixels labeled as 1 within ROI, while pixels valued as 0 if they are not within ROI. Secondly, we can then obtain the estimated values of the normalized GLCM. There are a total number of 22 textural features that could be computed from the GLCM, such as described in [46] with the detailed GLCM explanation.
2) In the spatial domain, which is the most common representation in the computer world, an image is comprised of many pixels and can easily be stored by a 2D matrix. In addition to the representation in the spatial domain, an image can also be represented in the frequency domain through the well-known spread spectrum approach like the Discrete Wavelet Transform (DWT). Spectrum analysis is another form of textural analysis in which the direction and the wavelength are considered. Based on the research of [8, 13, 46], the feature set in this study focused on a two-dimensional scaling wavelet that is a product of two one-dimensional functions.
3) Gaussian filter is very helpful in minimizing the noise in the image using the process of convolution. It is a class of low-pass filters, which are based on Gaussian probability distribution function [13]. Convolution is the sum of the whole matrix, by multiplying matrix filter with neighboring extension of the point (x, y)on the image. Gaussian 2D distribution is defined in the Appendix Table 12.
4) The Laplacian of Gaussian (LoG) filter is the second derivative of Gaussian filter [13]. This filter combines noise reduction and only responds to changes in image gradients, as they are in neighborhoods of pixels, and also respond to variations over small regions. The filter is based on partial derivatives of Gaussian Kernel and commonly used derivative filters including the Laplacian filter [39]. The LoG filter smoothes the image with a Gaussian low-pass filtering, followed by the high-pass Laplacian filtering.
5) The unsharp filter is a sharpening operator which enhances edges (and other high frequency components in an image) via a procedure which subtracts an unsharp, or smoothed version of an image from the original image. The unsharp filtering technique is commonly used in the photographic and printing industries for crispening edges. We therefore used Unsharp filters with the high-frequency Laplacian filters to retrieve the information [39, 48].

6) The Wiener filtering is applied to images with a cascade implementation of combination between the noise smoothing and inverse filtering which can remove noise more effectively. It is an adaptive linear filter that works on local variant characteristics of an image. This filter can make image areas smoother than most visible noise, but retains areas where details are very visible while noise is less visible [48]. It is a linear estimation of the original image and has been widely used in image denoising. Furthermore, it has good performances on removing the Gaussian white noise.
7) Gabor filter is a two-dimensional filter which has the Gaussian kernel function modulated by a complex sinusoidal plane wave and has several advantages such as invariance to illumination, rotation, scale, and translation [14]. In addition, it is designed for different dilation and rotation operation, with the filter bank to analyze the texture of images.
8) The co-occurrence matrix and texture features are the most popular second-order statistical features that are introduced by RM Haralick in 1973. He extracted different statistical features known as Haralick texture features, using GLCMs by computing various statistical properties that were used to construct the matrix obtained by employing the directions 0^0, 45^0, 90^0, and 135^0 [15]. In a microscopic image, the spatial grayscale level reliance matrix at the direction, and the spatial distance such as GLCM feature filter where (i, j) indicates the spatial location of image. *Glcm(n, m)* means the number of occurrences of pixels with graylevels n and m respectively, with a separation of (dr,dc) pixels.

2.4.3 The fractal features

The segmentation-based fractal texture analysis (SFTA) features are implemented in image analysis with the content similar texture. This feature is built based on fractal dimension for gray-scale images depicting object and structure boundary segmented images. SFTA not only computes the fractal dimension of any grayscale images but also measures the roughness of the images with different textures. The extraction algorithm consists in decomposing the input images into a set of binary images from which the fractal dimensions of the resulting regions are computed, in order to describe the segmented texture patterns. The SFTA feature vector - $\Delta(x, y)$ is defined in the Appendix Table 12. Costa [9] and Bekhti [2] implemented the threshold set of the binary (n_t), corresponding to the maximum possible gray level in $I(x, y)$ images which yield 48 features.

2.5 Support vector machine

The texture classification is essentially the problem of classifying pixels for images according to the textural cues. SVM concepts can be simply explained as an attempt to find the best hyperplane which serves as a separator among classes in the input space. The best separation among hyperplanes can be found by measuring the margin hyperplanes, and searching for the maximum points [24]. The SVM generates a model based on the training data, and can predict the target values of the test data given only the test data attributes [18]. SVM can be applied by combining extractor features to obtain the best result, in comparison among feature extractions for a multi texture classification problem. We chose the radial basis function, the (RBF)-based kernel

function, to build the classifier for our study as defined in Eq. (1), where γ is a parameter that sets the "spread" of the kernel.

$$K_{RBF}(x_i, x_j) = \exp\left(-\gamma \left\| x_i, x_j \right\|^2\right), \gamma > 0 \quad (1)$$

We classified the image that has been extracted in different features using SVM utilities to obtain the optimized parameters. Optimal kernel parameter for C and γ were obtained by a coarse grid search in the parameter space within the interval C, and a mapping ϕ is considered to transform the original data space into another featured space, as suggested in [18].

3 The proposed method

Inspired by the researches [26, 50], we propose the technique based on the machine learning method by examining the microscopic images from not only text, but also the picture images.

3.1 Experimental steps

The diagram in Fig. 3 illustrates the identifying procedures which can be divided into three stages, such as digitizing documents, feature extraction and classification:

(1) *Digitizing documents*: First of all, we prepare text and image documents in Microsoft Word format. For example, font-type Arial with the font size 10 pt. for characters are printed. In this study, several printed characters are identified like English character "e", Chinese character "永", Arabic character "ج", and Japanese character "シ". We also investigated the image documents (i.e. Lena, Pepper, and Baboon) with 512 × 512 pixel format. Secondly, we print the document by using 12 different printers as shown in Table 2. After all the documents have been printed, the third step is to digitize the documents by using a reflected microscope Olympus CX 41 with different magnification. Each microscopic image is then cropped into patches that represent the documents based on the different printer sources. At this stage, two different document types are identified, i.e. microscopic images for text and the image documents.

(2) *Feature extraction*: Extracting the grayscale documents by the proposed filters such as LBP, GLCM, DWT, Gaussian. LoG, Unsharp, Wiener, Gabor, Haralick, and fractal filters. In this step, we extract the document separately based on text and image documents. All of the microscopic images that have been cropped in patches, we obtain each character or image based on the printer sources. For example, the character "e" that originates from each type of printer, we take out at least 1200 images for each printer by using different feature filters into numeric values.

(3) *Printer classification*: The last step of our approach is to classify the features found thourgh each feature filter, by using SVM trained model to identify the printed sources from 12 different printers. The extracted images that have been in the numeric value are then inserted into the MySQL database. The database containing different schemes and queries based on printed document types will be evaluated. Afterwards, we classify them by using SVM in the Java environment (Eclipse Indigo) with connection to the database. In this study, we adopted the settings by previous researchers Mikkilineni [30] and Tsai [46], who applied 500 images for training and 300 images for testing respectively. To

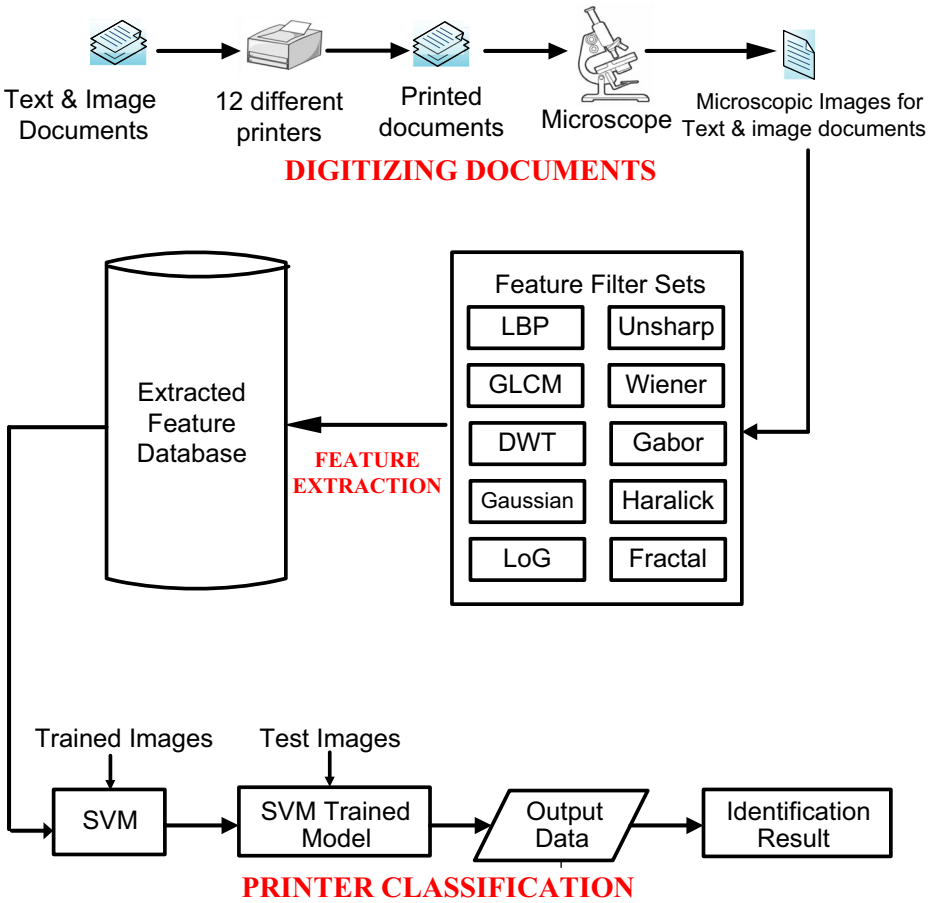


Fig. 3 The pipeline of printed source identification

implement the mixed data, we used different settings of the training and test data. For example, when we mixed the data from e_1 & e_2, (English character "e" printed in different font or size will be noted as different data), 1000 images will be needed as the raining data and another 600 images for the testing data.

3.2 Experimental setup

To create a consistent analysis for fair comparison, issues regarding the experimental setup such as the paper sample, the software platform, and the appropriate microscopes for digitization, all need to be considered before conducting the different experiments.

3.2.1 Paper sample

Mikkilineni, et al. [30] had tested different paper types for the character 'e', and thus obtained various classification results using different paper types. On the other hand, they acquired high

Table 2 Printer brand and models

No	Brand	Model
1	Avision	AM/MF 3000
2	HP	LaserJet Pro 200 Color P. M251nw
3	HP	LaserJet Pro 500 MFP M570dn
4	HP	Color LaserJet CP3525
5	HP	LaserJet Pro CP1025
6	HP	LaserJet 4300
7	HP	LaserJet 4200dtn
8	HP	LaserJet M1132 MFP
9	HP	LaserJet Pro 400 MFP M425dn
10	HP	LaserJet M1522nf
11	HP	LaserJet Pro M1536dnf
12	OKI	C5900

accuracy classification when both the training and testing sets use the same paper. By using micro-scale imaging, the different types of paper have varied structure and surface [43] which could influence the quality of the printed documents and images [38]. Since the surface of a sheet of paper is not perfectly flat, a printer will produce different images as a result of the different orientation of the pages. This means that the surface roughness of the paper, the absorption of the ink and fountain solution, will cause the non-uniformity of ink transfer [22]. Furthermore, the surface roughness of coated and uncoated papers will influence the occurrence of the print mottles [20]. Thus, the selection of the paper will have a strong impact on the printing properties, including a huge variety of characteristics such as its particles, shape, and chemical composition [21].

Similarly, Ferreira et al. [12] used white paper of the same brand to explore and compare printer attributes on source identification. The influence of the paper texture is treated irrelevantly in this study, and the same brand paper in weight 80 g of Paperone, all white color and A4 size (210 × 297 mm), is adopted in this study for all tested printers.

In this study, the printed document in microscopic image is equally divided into a number of patches. As shown in Fig. 4, given a microscopic image I with width w and height h, a patch refers to an area where it is defined by the top (p_x) to the left point (p_y) and size s, $1 \le s \le \min(w, h)$. Next, all the patches are indirectly specified by the sampling stepτ. Obviously, if τ equals s, a microscopic image would be segmented into grids without overlap. Under those constraints, the border patches and paper region that are moderately outside the scope of ROI

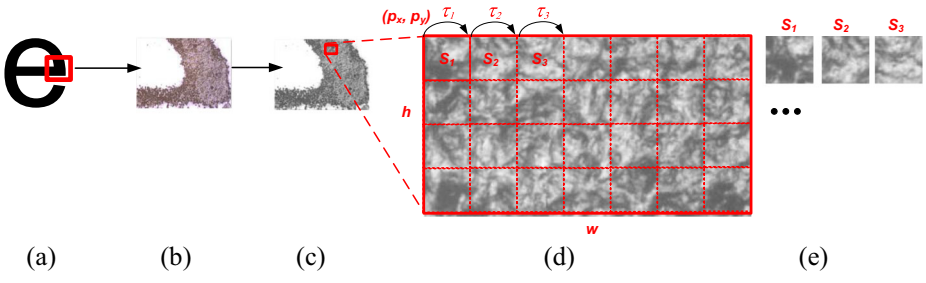


Fig. 4 The technique of acquiring microscopic samples (**a**) character "e", (**b**) microscopic image, (**c**) grayscale image, (**d**) image patches, and (**e**) image samples

are avoided because they contain no sufficient information. From Fig. 6, those patches clearly show the texture information from different printers.

3.2.2 Applying a suitable optical microscope

In order to choose the appropriate microscope in this study, a pre-experiment test is performed by using two different optical microscope types i.e., the transmitted light microscope (TLM) and the reflected light microscope (RLM). For this purpose, Arabic character "ج" (*jīm*) with Arial 10 point is scanned by 20× objectives lens using TLM Zeiss Axio Imager 2 and RLM Olympus CX 41. The images produced by both optical microscopes are shown in Fig. 5. The images obtained from both microscopes have dissimilar textures and highly contrasted patterns. All of the images from TLM are dark and black in general. In contrast, the RLM's texture shows clear, intact particles and areas of molten particles. Thus, the microscopic images can be obviously identified by the toner particles based on the color, the size and the shape. Furthermore, the structural differences between the paper surfaces and the toner particles can be clearly distinguished from each other. Overall for this study, the most suitable microscope observing the microscopic image pattern turns out to be the RLM type.

Generally, the magnification system with microscope helps to increase the resolution of the image. For example, 10× of objective lens and 10× of subjective lens yield the magnification 100×. Digital image magnification is calculated by the following formula: CCD chip size / (objective magnification x coupler magnification) [11, 29]. For this study, Olympus CX 41 digital camera is adopted which has the resolution 3.1 M pixels and CCD Chip is 2048 × 1536 pixels. Each image generated by the microscope has 1600 × 1200 pixels. Hence, as shown in Fig. 1, the maximum of digital magnification is 1600 / (2048 μm / 10) = 7.81 pixels/μm. All microscopic images implemented in this study have the same pixel size for the text and image document.

3.2.3 The software and tools platform

In this study, the digital camera Lumenera's INFINITY 1 with 3.1 M pixel captures images of the optical microscope Olympus CX 41. The software named the INFINITY ANALYZE

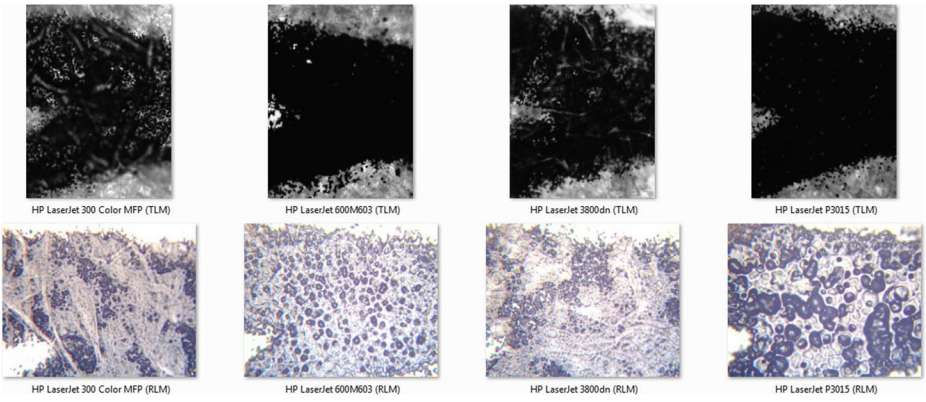


Fig. 5 The comparison of images which are obtained from two different microscopes. The first row images are from TLM and the second row images are from RLM

software is used for advanced camera control, image processing, measurement and annotation. Afterwards, Netbean IDE 8.1 is adopted to acquire microscopic images by cropping the region of interest. Next, feature extraction and SVM implementation is performed by Matlab R2013a with Eclipse SDK under Java environment.

4 Experimental results and discussion

In this study, we have conducted a series of experiments to identify the printer source from microscopic images of text and image documents. Different situation has been considered such as mixed text and image data, different data sizes, different types of microscopes and others. The following explains the experimental results:

4.1 Experiment I: Classification for text documents

For the text documents, not only English characters, but also Chinese, Arabic, Japanese characters are performed for comparison, for example, the English character "e", Chinese character "永", Arabic character "ع", and Japanese character "シ". All microscopic images for the text documents are in JPG file format with the dimension of 1600 × 1200 and file size 2.73 KB. Next, all the images are converted into grayscale in bitmap file format (BMP). After digitizing the documents, the images are cropped by using software Netbean IDE 8.0. Fig. 4 shows the cropping method for the acquired image patches [26] of the text document "e", with pixel size 43 × 43 and the file size 2.90 kilobytes in each microscopic image.

The English character "e" is a common character widely tested by researches [12, 30, 46]. As shown in Fig. 6, the image samples at the same location for character "e" from 12 different printers are illustrated for comparison. All of the images of the characters "e" are taken by applying 10× objective lens and 10× ocular lens, and the original microscopic images output is 1600 × 1200 pixel size. There are fundamental differences in the textures among the 12 images, but some of them also have similar textures. For instance, the images of Fig. 6(j) from HP LaserJet 4200dtn and Fig. 6(k) from HP LaserJet 4300 are similar. According to HP

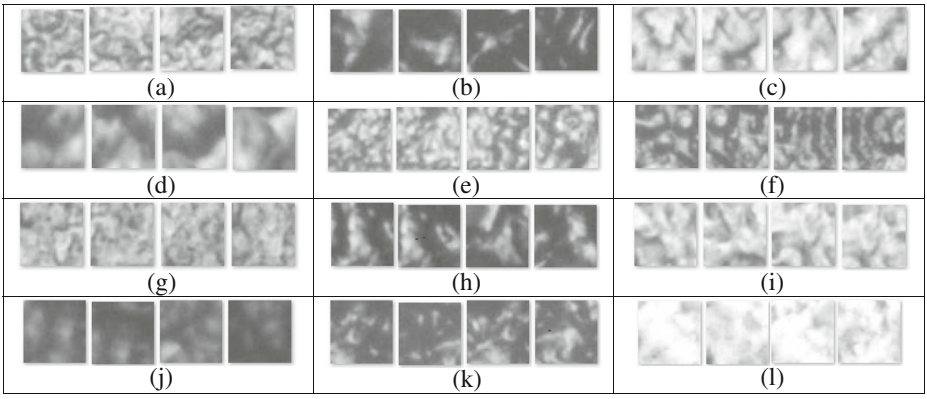


Fig. 6 image samples for "e" from 12 different printers in 43 × 43 pixel size. **a** HP LaserJet Pro 200 Color P. M251nw, (**b**) HP LaserJet Pro 400 MFP M425dn, (**c**) LaserJet Pro 500 MFP M570dn, (d) HP LaserJet M1132 MFP (**e**) HP LaserJet Pro CP1025 (**f**) HP LaserJet M1522nf, (**g**) HP LaserJet Pro M1536dnf, (**h**) Avision AM/MF 3000, (**i**) HP Color LaserJet CP3525, (**j**) HP LaserJet 4200dtn, (**k**) HP LaserJet 4300, and (**l**) OKI C5900

LaserJet user guide [17], both printers adopt FastRes 1200 to produce 1200-dpi print quality for fast, high-quality printing of business text and graphics. In addition, both printers also apply ProRes 1200 to produce 1200-dpi printing for the best quality of text and graphic images. ProRes and FastRes are the native 1200 dpi print engines for both printers. Consequently, the images from both printer models generate similar usage fingerprint from the printers.

Other microscopic images in Fig. 7 are produced by different laser printers which show clear particles, shapes, and bubbles from the microstructure of printed documents captured by RLM optical microscope. Visually, the toner particles printed on paper are generally spherical, whereas the particles themselves are actually angular and irregularly shaped. For illustration purposes, enlarged images with same resolution from three different printers are being compared. Since their printing engines are totally different, a specific characteristic pattern is yielded from each printer and resulting visual differences. Notably, the same printer produces similar texture and coloring in the patterns.

To classify the printer source by using the proposed approach, the accuracy rate based on character "e" is first investigated from previous research [12, 30, 46]. In this study, we analyze not only character"e" but also mixed characters e_1, e_2, and e_3 which have different font and size. Before classification, all microscopic images were firstly extracted by using 10 feature extraction sets, such as mentioned in section 2. Then, the whole extracted images with numeric characteristic value files are exported to MySQL database and classified based on filter sets, by using SVM in the Eclipse of Java environment. The accuracy rates to predict the printer source for character "e" are tabulated for comparison.

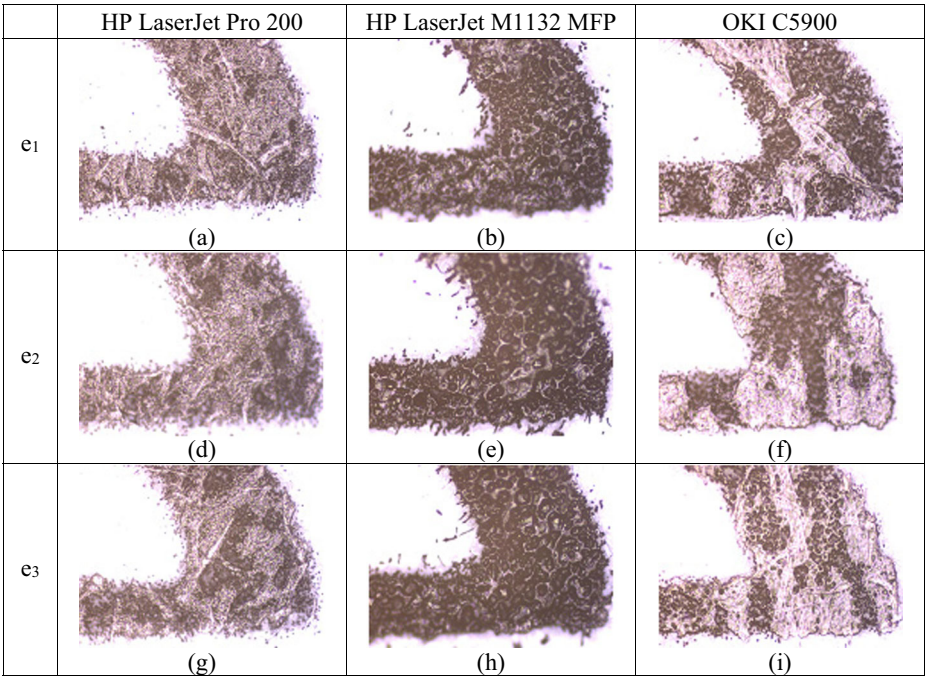


Fig. 7 e_1, e_2, and e_3 are English character "e" printed in different font and size with 160 × 120 pixel size from different printer. Figure (**a**), (**d**), and (**g**) are from Printer HP LaserJet Pro 200 Color P. M251nw. Figure (**b**), (**e**), and (**h**) are from Printer HP LaserJet M1132 MFP. Figure (**c**), (**f**), and (**i**) are from Printer OKI C5900

The steps of proposed approaches are listed below:

(1) 10 sets of images from the microscopic image database of 12 printer sources are randomly generated. In each set, there are 500 images which are selected from each printer as training data and another 300 images for test data. 10 sets of feature filters are then applied for characteristic extraction.
(2) Apply the SVM engine to build the prediction models using 10 sets of feature filter.
(3) Feed the test image subsets to the corresponding model trained in step 2 for the printer source prediction.
(4) Repeat step 1 through 3 ten times to obtain the predicted results.

Table 3 tabulates the accuracy rate in the three different printed characters "e" from their microscopic documents, by using ten different set of filters in the same character. It can be seen that DWT feature set has the lowest accuracy prediction for microscopic images character "e" with the average percentage below 80%. The Haralick and Gaussian feature sets have slightly higher accuracy ratios than DWT. The other feature sets i.e., Gabor, Wiener, GLCM, LoG, Unsharp, and SFTA also have good results for printer source identification. Among those feature sets, LBP feature set achieves the highest accuracy ratios. For example, its character e_1 is at 99.89%, and characters e_2 and e_3 are also at comparable high accuracy.

A line graph shown in Fig. 8, illustrates the accuracy rates for different "e"s based on ten different set of feature filter. It is apparent that the best set of filters to predict printed source using microscopic images is LBP filters. The GLCM, UnSharp, Wiener, Gabor, and SFTA filters also have high capability for identification. On the other hand, the DWT, Gaussian, and Haralick have the lowest percentages in the table when the mixed data are examined in the experiment.

4.1.1 Character "*e*", "ح", "シ", and "永"

To observe and compare the text from different printers and different alphabets, character "e" (English), character "ح" (Arabic), character "シ" (Japanese) and character "永" (Chinese) are examined. The characters are taken from the same microscope with the same magnification. As shown in Fig. 9, the microscopic images which are printed from Printer HP Color LaserJet CP3525 are illustrated for comparison. Fig. 9(a) is English character "e"; Fig. 9(b) is Arabic

Table 3 The accuracy rates for character "e"

Filter	e_1 (%)	e_2 (%)	e_3 (%)	Average e_1, e_2 & e_3	Mixed e_1 & e_2	Mixed e_1, e_2 & e_3 e_1,e_2,e_3
LBP	99.89	99.87	99.94	99.90	99.77	99.58
GLCM	94.71	95.18	94.58	94.82	87.94	80.56
DWT	77.24	75.90	82.74	78.63	66.56	63.37
Gaussian	86.29	83.65	83.85	84.60	75.81	68.41
LoG	93.68	93.58	95.98	94.41	85.88	82.58
Unsharp	94.16	94.61	95.66	94.81	88.41	83.36
Wiener	96.39	97.73	98.12	97.41	95.48	94.45
Gabor	96.58	98.11	98.87	97.85	95.49	93.34
Haralick	84.62	87.24	85.07	85.64	75.38	68.23
SFTA	91.27	91.53	93.52	92.11	84.09	79.75

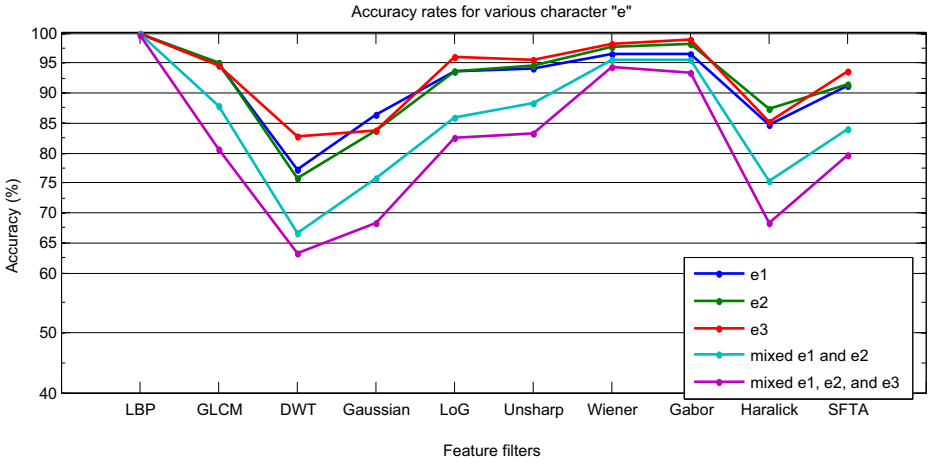


Fig. 8 A line graph of the percentage of accuracy rate for different "e"s based on ten feature filter sets

character "ح"; Fig. 9(c) is Japanese character "シ"; and Fig. 9(d) is Chinese character "永". The printed characters are from the same printer and the particle shape such as bubbles and dots are similarly grouped to form the letters.

The printer source identification procedure is similar to the previous section which is performed on different alphabet. As shown in Table 4, the accuracy prediction rate for characters "e", "ح", "シ", and "永", have stable percentages on each feature set of filters. For example, the accuracy prediction for character "ح "is 74.61% for DWT filter, 86.36% for Haralick features, for 86.76% for Gaussian filters and other filters (GLCM, LoG, Unsharp, Wiener, Gabor and LBP) are above 90%. Overall, LBP features are superior in analyzing microscopic image textures for the different alphabet.

4.1.2 Data mixed from character "e", "ح", "シ", and "永"

To implement the mixed data, we used different settings of training and test data. For example, when we mixed the data from e_1 & e_2, there are 1000 images which are selected from each printer as training data and another 600 images for test data. As well, there are 1500 images as training data and 900 images as test data for mixed data from e_1, e_2, & e_3 . For mixed data, the procedure of the steps can be shown in Fig. 10. According to Table 4 and Fig. 11, the accuracy rate of source identification using mixed texts is still high, but the average performance

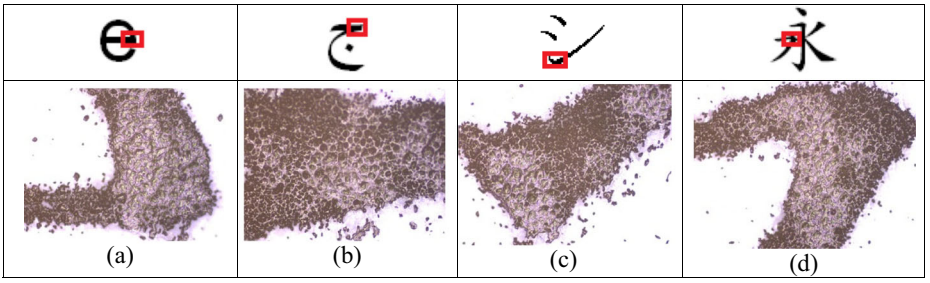


Fig. 9 Microscopic images which are printed from Printer HP Color LaserJet CP3525. Figure (**a**) is English character "e", (**b**) is Arabic character "ح", (**c**) is Japanese character "シ", and (**d**) Chinese character "永"

Table 4 The accuracy rates for different alphabet

Filter	English e	Arabic ج	Japanese シ	Chinese 永	Mixed data e, ج, シ, and 永
LBP	99.89	99.92	99.91	99.89	99.39
GLCM	94.71	95.13	94.07	92.68	79.45
DWT	77.24	74.61	71.61	70.20	53.56
Gaussian	86.29	86.76	86.34	83.14	66.60
LoG	93.68	94.68	94.84	93.19	81.40
Unsharp	94.16	95.02	94.45	93.66	80.27
Wiener	96.39	96.53	96.13	95.87	91.04
Gabor	96.58	96.06	97.28	97.07	89.04
Haralick	84.62	86.36	82.27	80.45	64.75
SFTA	91.27	90.96	89.41	93.10	76.49

declined when compared to use only one character. From Table 4, the LBP features score above 99% and the DWT is under 75%. In addition, the other feature filter sets which score accuracy rates above 90% are GLCM, LoG, Unsharp, Wiener, Gabor and SFTA. Thus, using LBP features still achieve the best results in this experiment.

4.2 Experiment II: Classification for image documents

Since image documents are quite commonly accepted for circulation, it is also necessary to identify the printer source based on image documents. Therefore, the Lena, Peppers, and Baboon images are tested in this experiment. Unlike the alphabets, the region of interest in the images is universally important across the image. Therefore, a given region for each image is selected as shown in Fig. 12. Due to the limited space in this journal paper, we do not show all the microscopic images of image documents from 12 different printers.

Basically, a printer of the same brand and type produces similar textures. Apparently, images differ in their density and the spread of toner printed on the paper. Fig. 13 demonstrates a technique for acquiring microscopic samples in five different areas. The original image file type was bitmap with a dimension 1600 × 1200, and the file size is 2.73 MB. We then cropped the microscopic file into 43 × 43 dimensions, with 2.90 kilobyte for the size. All remaining cropping work was similarly implemented as explained in Sec. 3.2. We obtained 1200 image samples in each printer brand. Thus, the number of sample in each image document is 14,400

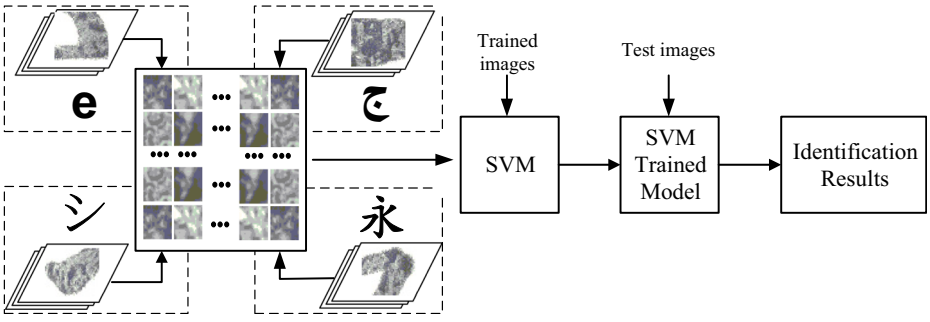


Fig. 10 The pipeline of data mixed from different character

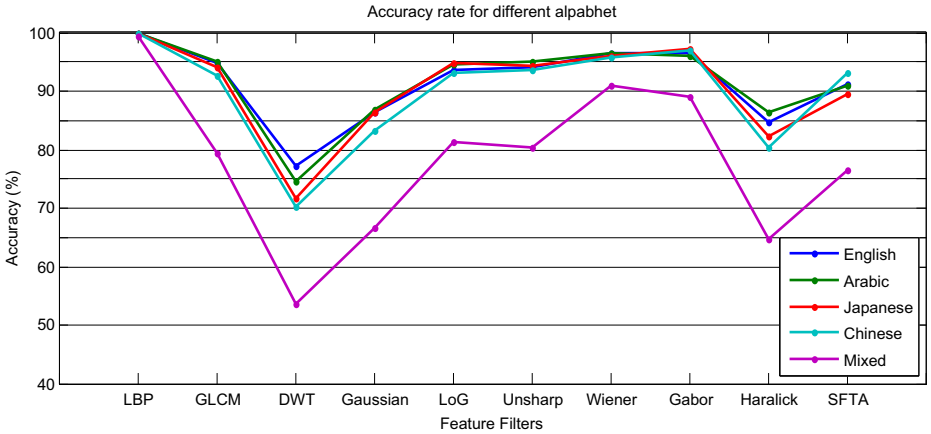


Fig. 11 A line graph of the percent accuracy rate for different alphabet based on ten feature filter sets

images from 12 printers. The steps of the experiment and the number of the test data, and the training data are as follows:

(1) 10 sets of images from microscopic image database of 12 printer sources are randomly generated for image documents like Lena. In each set, there are 500 images which are

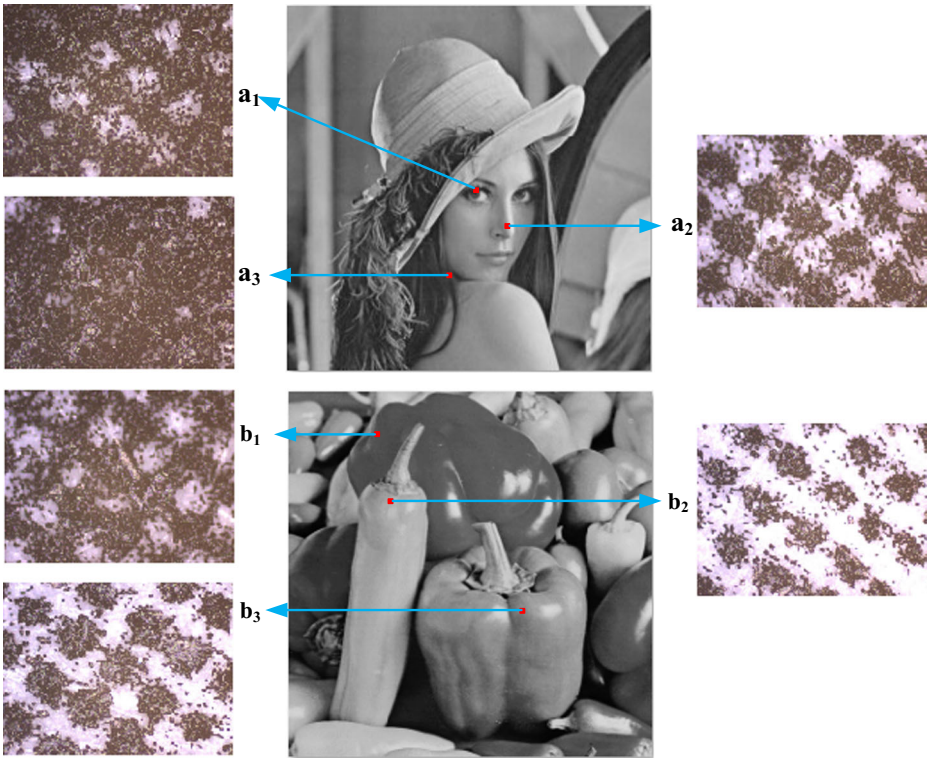
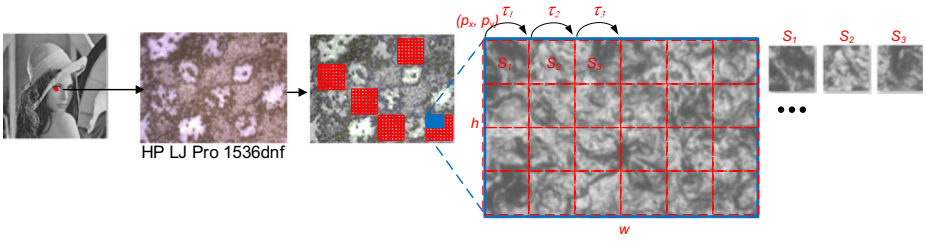


Fig. 12 "■"capturing microscopic images from Printer HP LaserJet Pro 400 MFP M425dn in 239 × 179 pixel size



(a) (b) (c) (d) (e)

Fig. 13 The technique of acquiring microscopic samples (**a**) character "e", (**b**) microscopic image, (**c**) grayscale image, (**d**) image patches, and (**e**) image samples

selected from each printer as training data, and another 300 images for test data. The 10 set filters (LBP, GLCM, DWT, Gaussian, LoG, Unsharp, Wiener, Gabor, Haralick, and SFTA features) are then calculated.

(2) Apply the SVM engine to build the prediction models, using all features.
(3) Feed the test image subsets to the corresponding model trained in step 2 for the printer source prediction.
(4) Repeat step 1 through 3 up to ten times to obtain the predicted results
(5) Change the data with each image document for Peppers or Baboon, and then apply step 1 to 4.
(6) Mix the image data among the image documents, i.e., mixing Lena, Peppers, and Baboon. We trained 3600 images as feature vectors from all images from every printer. In each set, there are 1500 images which are selected from each printer as training data, and another 900 images for test data.

Table 5 provides summary statistics for the accuracy prediction for printer source to identify image documents (Lena, Peppers and Baboon) among different filters. The LBP feature still achieved the highest prediction (96.69%), followed by Gabor and Wiener above 80% accuracy rates. Meanwhile, the LoG and SFTA filters score above 70% with the UnSharp and GLCM under 70%. It is interesting to see the image forensics when the text and image documents are mixed. For the text and image mixed documents, the data from e_1 and $Lena_1$ are studied. Specifically, 2400 microscopic images (1200 images from text document and also 1200 images from image document) from each printer brand are collected. The procedures in this

Table 5 The accuracy prediction by mixing among image documents

Filter	Lena (L) (%)	Peppers (P) (%)	Baboon(B) (%)	Mixed L, P & B (%)
LBP	99.54	98.47	98.69	96.69
GLCM	85.89	84.76	81.01	61.27
DWT	72.5	75.73	67.95	53.21
Gaussian	77.86	74.96	75.73	53.17
LoG	89.72	91.5	87.27	73.02
Unsharp	89.34	92.03	87.07	66.57
Wiener	94.19	91.92	84.86	81.50
Gabor	96.71	93.29	93.79	85.70
Haralick	75.45	72.84	73.49	51.17
SFTA	88.47	88.54	87.73	73.85

Table 6 The accuracy prediction by mixing text and image documents

Filter	e_1 (%)	Lena (%)	Mixed data Text (e_1) and Image (L) (%)
LBP	99.89	99.54	99.39
GLCM	94.71	85.89	80.88
DWT	77.24	72.5	66.08
Gaussian	86.29	77.86	72.13
LoG	93.68	89.72	86.06
Unsharp	94.16	89.34	83.98
Wiener	96.39	94.19	93.66
Gabor	96.58	96.71	94.78
Haralick	84.62	75.45	70.08
SFTA	91.27	88.47	84.32

experiment are the same as those in the mixed characters e_1 and e_2. Table 6 tabulates the results of the experiment by mixing text and image documents. Overall, the ten set of filters which are applied in feature extraction to classify the printer brand contribute different results. The accuracy percentage in the text document is higher than in the image document. For example, the accuracy rates of GLCM, LoG, Unsharp and Haralick features for the text documents are above 90% but are under 90% for the image documents. Similarly, the accuracy rates of the filters using Gaussian and Haralick filters are under 80% for images, but are above 80% for text documents. Apparently, LBP filters still achieve the highest accuracy rates in this experiment, even when text and image documents are mixed.

4.3 Experiment III: Classification in different data sizes

Referring to the previous research [46], the bigger size sample data can provide more information achieving higher performance accuracy rates in the printer source identification. For this reason, samples of different data size from microscopic images are also studied. With the three different data size for data "e" with the pixel 51 × 51, 59 × 59, and 67 × 67, Table 7 and Fig. 14 demonstrate the identification accuracy rate comparison using microscopic images in different data size. Overall, the numerical figures from the table indicate that the larger pixel sizes generally yield higher accuracy rates. From the experimental statistics, the average time to process 14,400 images for 12 printers using 43 × 43 pixels takes 64 min, but 96 min are needed when 67 × 67 pixels are applied. Thus, it is apparent that applying higher pixel size of

Table 7 The accuracy rates for character "e" in different pixel size

Filter	e (51 × 51) (%)	e (59 × 59)(%)	e(67 × 67)(%)
LBP	99.93	99.98	99.98
GLCM	94.33	98.01	97.32
DWT	82.06	81.19	84.51
Gaussian	85.57	90.01	94.59
LoG	91.45	94.87	95.49
Unsharp	94.08	96.98	97.32
Wiener	95.32	97.48	98.60
Gabor	96.43	97.24	98.03
Haralick	84.01	89.11	93.28
SFTA	91.05	95.31	97.24
Average	91.42	94.02	95.64

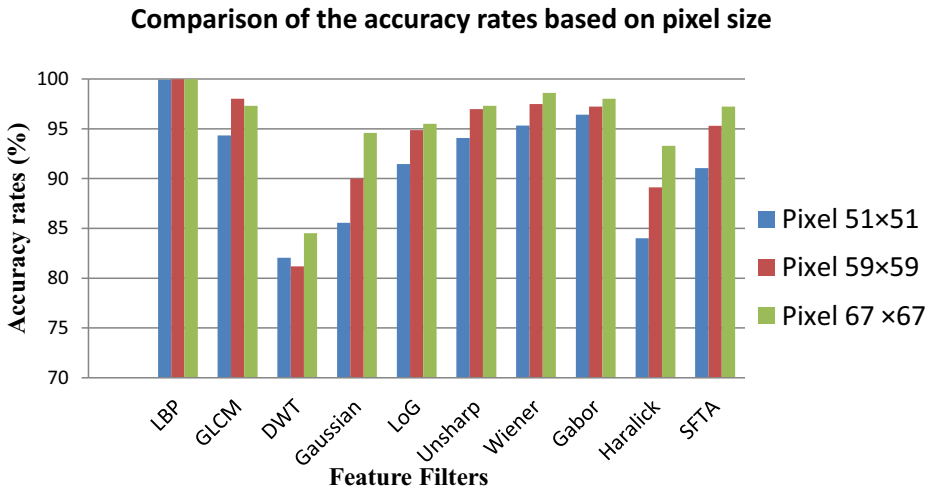


Fig. 14 comparison of microscopic images in different data size

microscopic images requires larger computation cost in order to complete the whole procedures.

4.4 Discussion

Based on the experiments in Sec. 4.3, 10 feature sets of filters and SVM classification are utilized to identify the printed source from either texts or images, and even both. After reviewing and comparing the statistics among feature sets, some findings are summarized as the following:

1) The benefits of acquiring detailed information using micro-scale imaging

The optical microscope with appropriate magnification can generate images with detailed texture information. With the help of the microscope, the intrinsic information regarding the shape, color, texture and roughness of a printer toner is used for the experiment samples

Table 8 Printed document source identification comparison based on text and Image document

	Research	Printed document	The method	Accuracy rate (%)
Text document	[12]	Character “e”	LBP	90.20
	[31]	Character “e”	GLCM	93.00
	[12]	Character “e”	CTGF_MDMS	97.60
	This study	Character “e”		99.89
	[46]	Character “永”	GLCM & DWT	98.23
	This study	Character“永”		99.89
	[47]	Char. “ シ ”	GLCM, DWT, other spatial features	90.88
	This study	Char. “ シ ”		99.81
Image document	[12]	Image document	LBP	95.21
	[6]	Color Image	GLCM,DWT	99.34
	[12]	Image document	CTGF_GLCM_MDMS	98.47
	[23]	Image document	DFT	94.40
	This study	Image document		99.54

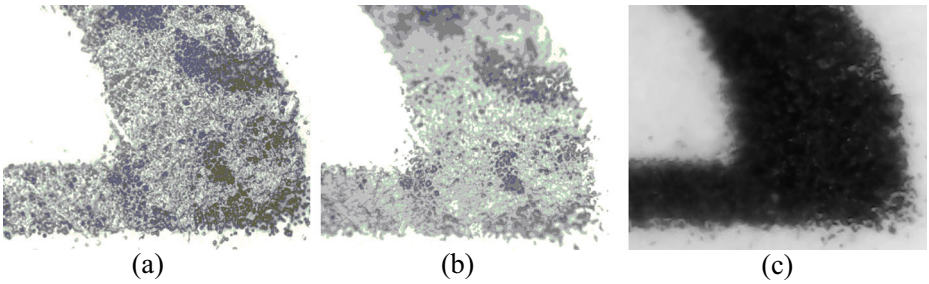


Fig. 15 HP Color LJ CP3525 for character "e" (a) Olympus CX41, (b) Olympus BX51M, and (c) USB CoollingTech

through the feature extraction. From the simulation comparison, the superior accuracy rates using microscopic images outperform the results using scanner devices [12, 31, 46]. From Table 8, the proposed technique using microscopic images achieves the highest accuracy rates for printer source identification which shows significant better results than the others.

2) The evaluation using different microscope

To evaluate the consistency results in this study, we also applied different reflected light microscopes, i.e., Olympus CX 41, BX 51 M and USB microscope. Similarly, we evaluate the character "e" that is captured by different lens sizes and magnifications. We conducted an experiment to justify the proposed approach based on the images in different microscopes. Visually, there are different results among the images as shown in Fig. 15 and Fig. 16. Both microscopes (CX 41and BX 51 M) illustrate clear speckles of texture pattern where the USB microscope can not achieve. As shown in Table 9, the statistics indicate that the average accuracy rates when using the USB microscope is only 60.72%. In contrast, the Microscope CX 41 and BX 51 M that have higher resolution and illumination systems can score above 90% accuracy. The results demonstrate that low cost USB microscope used in [41, 50] is

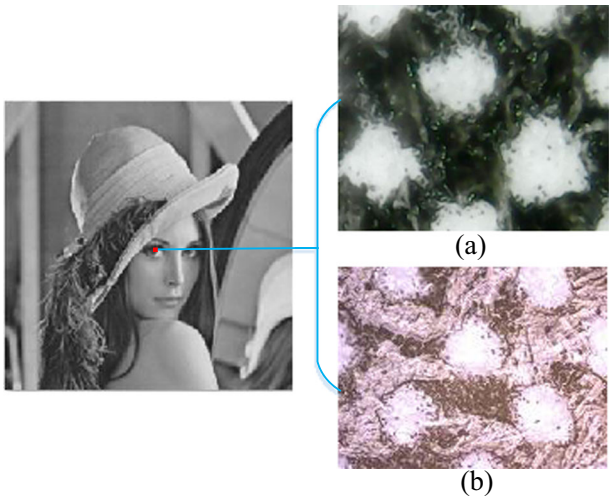


Fig. 16 Patches of Lena Image using (a) USB microscope (b) Olympus CX41 microscope

Table 9 The accuracy rates for character "e" in different microscope

Filter	Microscope CX 41(%)	Microscope BX 51 M(%)	USB Microscope (%)
LBP	99.89	99.97	79.98
GLCM	94.71	96.87	49.06
DWT	77.24	89.84	64.73
Gaussian	86.29	90.63	44.46
LoG	93.68	96.88	61.93
Unsharp	94.16	96.99	61.32
Wiener	96.39	98.31	70.97
Gabor	96.58	99.15	78.17
Haralick	84.62	88.40	44.18
SFTA	91.27	95.81	52.35
Average	91.48	95.29	60.72

suitable for paper authentication but not for the purpose of source identification forensics. On the other hand, the professional-grade microscopes like the CX 41 and BX 51 M, can achieve comparably better results.

3) The highest and the lowest accuracy rates in printed sources prediction

As shown in Table 10 and Table 11, the confusion matrixes from LBP filter and DWT filters are calculated by comparing the percentages of accuracy rate in predicting printer sources, corresponding to the prediction and the actual printer sources in the classification stage. For example, the accuracy rate in Printer Avision AM/MF 3000 identification at Table 10 is 5968/6000 = 0.9947 or 99.47%. The average of accuracy rate prediction by using LBP is above 99%. It means that the filter yields the highest accuracy rate to investigate printer sources. Conversely, by using DWT filters as shown in Table 11, each column of the confusion

Table 10 Confusion matrix LBP Filter which is the highest accuracy rate in the mixed document

Avg 99.39		Predicted (%)											
		1	2	3	4	5	6	7	8	9	10	11	12
Actual (%)	1	99.47	0.00	0.00	0.02	0.25	0.08	0.07	0.02	0.00	0.08	0.02	0.00
	2	0.07	99.53	0.03	0.02	0.02	0.10	0.00	0.03	0.05	0.07	0.03	0.05
	3	0.00	0.10	99.53	0.00	0.00	0.02	0.00	0.10	0.17	0.00	0.05	0.03
	4	0.05	0.05	0.00	99.87	0.02	0.00	0.00	0.00	0.00	0.02	0.00	0.00
	5	0.40	0.00	0.00	0.03	99.30	0.00	0.22	0.02	0.03	0.00	0.00	0.00
	6	0.17	0.15	0.05	0.00	0.02	98.92	0.10	0.05	0.18	0.18	0.07	0.12
	7	0.07	0.18	0.02	0.00	0.05	0.17	99.18	0.03	0.00	0.22	0.03	0.05
	8	0.00	0.22	0.05	0.00	0.03	0.02	0.03	99.35	0.15	0.00	0.15	0.00
	9	0.00	0.30	0.30	0.05	0.03	0.03	0.05	0.07	99.12	0.00	0.05	0.00
	10	0.12	0.13	0.00	0.00	0.30	0.15	0.05	0.02	0.03	99.18	0.00	0.02
	11	0.00	0.12	0.07	0.00	0.00	0.02	0.07	0.03	0.05	0.08	99.57	0.00
	12	0.02	0.05	0.03	0.00	0.02	0.10	0.05	0.03	0.00	0.02	0.02	99.67

The highlight number refers to percentage hits of accuracy rate for predicting printer sources.

(1) Avision AM/MF 3000, (2) HP LaserJet Pro 400 MFP M425dn, (3) HP LaserJet 4200dtn, (4) HP LaserJet 4300, (5) HP LaserJet Pro 500 MFP, (6) HP Color LaserJet CP3525, (7) HP LaserJet M1132 MFP, (8) HP LaserJet M1522nf, (9) HP LaserJet Pro CP1025, (10) HP LaserJet Pro M1536dnf, (11) LaserJet Pro 200 Color P. M251nw, and (12) OKI C5900

Table 11 Confusion matrix DWT Filter which is the lowest accuracy rate in the mixed document

Avg 66.08		Predicted (%)											
		1	2	3	4	5	6	7	8	9	10	11	12
Actual (%)	1	65.57	0.18	0.12	0.07	8.20	11.10	0.65	0.25	0.03	12.68	0.85	0.30
	2	0.92	73.20	3.38	1.32	0.77	7.07	0.45	3.82	1.25	0.68	2.43	4.72
	3	0.02	4.72	68.55	0.07	0.00	1.90	0.07	6.37	3.38	0.42	7.40	7.12
	4	0.03	2.93	0.07	89.95	0.30	0.05	0.18	3.48	2.72	0.27	0.00	0.02
	5	16.20	1.53	0.00	1.30	75.80	1.70	2.00	0.13	0.07	1.03	0.00	0.23
	6	9.97	3.02	0.75	0.00	1.58	61.13	0.73	0.78	0.22	15.50	4.38	1.93
	7	16.53	2.60	1.37	0.02	4.43	9.67	49.80	0.00	0.03	4.05	9.05	2.45
	8	0.10	6.93	9.23	2.63	0.58	1.42	0.25	65.55	11.38	1.10	0.75	0.07
	9	0.22	5.50	6.40	2.18	0.43	2.57	1.07	19.18	58.68	1.37	1.12	1.28
	10	11.02	0.70	1.77	0.13	4.10	18.73	2.48	0.40	0.27	54.02	4.98	1.40
	11	1.18	1.85	11.38	0.00	0.03	8.27	6.15	0.20	0.05	4.02	65.07	1.80
	12	1.07	8.77	10.68	0.03	1.25	7.23	0.35	0.65	0.45	0.90	2.97	65.65

The highlight number refers to percentage hits of accuracy rate for predicting printer sources.

(1) Avision AM/MF 3000, (2) HP LaserJet Pro 400 MFP M425dn, (3) HP LaserJet 4200dtn, (4) HP LaserJet 4300, (5) HP LaserJet Pro 500 MFP, (6) HP Color LaserJet CP3525, (7) HP LaserJet M1132 MFP, (8) HP LaserJet M1522nf, (9) HP LaserJet Pro CP1025, (10) HP LaserJet Pro M1536dnf, (11) LaserJet Pro 200 Color P. M251nw, and (12) OKI C5900

matrix representing percentages of accuracy rate prediction has lower values than those in those in Table 10. Through systematic analysis, the best set of features can be discovered and identified for digital forensics accordingly.

4) Selecting and determining the appropriate samples

Selecting and determining the appropriate samples will affect the accuracy rate for printer source classification. In principle, printer toner that spread on the paper is not entirely uniform through the optical microscope. Therefore, an investigator has to be observant and alert in order to verify and determine the useful samples from the printer.

5) Deciding the most important features

Currently, there are a total of 306 features from the ten feature set. To search for the most important features and to reduce the evaluation time without the loss of accuracy, the adaptive feature selection algorithm could be implemented. According to [46], five feature selection algorithms in Java: SFFS, SBFS, plus-2-minus-1 (P2M1), plus-3-minus-2 (P3M2), and plus-4-minus-3(P4M3) could be implemented. The number of the chosen features is decided based on the accuracy rate for all 306 features. Currently, the selection process is not converged due to the huge number of features. Machine learning approach like CNN may be applied to reduce the total processing time for the feature selections.

5 Conclusion and future research

Most of the researches for printed forensics include using the scanner tools as the digitizing device. The work presented in this paper proposes the technique which analyzes the microscopic images of printed character and image documents for source identification. The

proposed technique utilizes SVM based classification approach with different feature sets such as LBP, GLCM, DWT, Gaussian, LoG, Unsharp, Wiener, Gabor, Haralick, and SFTA features. According to the experimental results, using microscope to magnify the intrinsic signature during the printing process improves the overall identification performance. The identification accuracy rate can score as high as 99.89% when using LBP features to examine the character "e", which is the highest rate compared with other feature set. Different language alphabet is also examined in this study, which also shows promising results from the experiments, and the identification accuracy rate achieves the highest score compared with other approaches. The superior performance demonstrates the efficacy of the forensic application using the proposed method.

It is apparent with the constraints of time, as well as the difficulties in accessing and using microscopes, certain preprocessing for microscopic images have their shortcomings, especially in retrieving reliable image samples during our investigation. Therefore, future research can be improved by getting samples automatically when the auto feed function is adopted for microscope. In addition, using different professional grade RLM microscopes during the investigation if resources are available, may bring further improvement.

Acknowledgments This work was partially supported by the National Science Council in Taiwan, Republic of China, under NSC104-2410-H-009-020-MY2.

Appendix: Formula of feature extraction

Table 12 Brief description of the formulas for ten feature filters are shown below

Feature filter	Image quality measures	Formula
LBP	$LBP_{P,R}(x_c, y_c)$ LBP features where P sampling points on a circle of R radius	$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s\left(g_p - g_c\right) 2^p s(x) = \{1, if x \geq 0;$ 0, otherwise. where P sampling points on a circle of R radius, are shown to form P, R and g_c corresponds to the grey value of the centre pixel, and g_p to the grey values of its neighbor p
GLCM	Region of interest R (ROI)	$R = \sum_{(i,j) \in ROI}^{1}$
	GLCM	$\text{GLCM}(i,j) = \frac{1}{\sum_{(i,j)} Img(i,j)} Img(i,j)$ where (i,j) indicates the spatial location of image. $Img(i,j)$ is the probability from location (i,j).
DWT	3 wavelet functions	$\Psi^{(H)}(x,y)$, $\Psi^{(V)}(x,y)$, and $\Psi^{(D)}(x,y)$, When the wavelete function is sparable by $f(x,y) = f_1(x)$, $f_2(y)$, then these functions rewritten to $\phi(x,y) = \phi(x), \phi(y)$ $\Psi^{(H)}(x,y) = \Psi(x), \phi(y)$ $\Psi^{(V)}(x,y) = \phi(x), \Psi(y)$ $\Psi^{(D)}(x,y) = \Psi(x), \Psi(y)$ where $\Psi^{(H)}(x,y)$, $\Psi^{(V)}(x,y)$, and $\Psi^{(D)}(x,y)$ are called horizontal, vertical, and diagonal wavelets
Gaussian	G $(x,y)$is Gaussian matrix element at position (x,y)	$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2\pi\sigma^2}}$ where G $(x,y)$is Gaussian matrix element at position (x,y), σ is the standard deviation.
LoG		$\text{Log}(x,y) = -\frac{1}{\pi\sigma^4}\left[1 - \frac{1}{\pi\sigma^4}\right] e^{-\frac{1}{\pi\sigma^4}}$

Table 12 (continued)

Feature filter	Image quality measures	Formula
	Log(x, y) is the high-frequency Laplacian filter	
Unsharp	$f_s(x,y)$ is the sharped imaged from unsharp mask	$f_s(x,y) = f(x,y) - \overline{f}(x,y)$
		where $\overline{f}(x,y)$ is a blured version of $f(x,y)$
Wiener	$H(u, v)$is function of Wiener filter	$G(x,y) = f(x,y) + n(x,y)$ $H(u,v) = \frac{P_f(u,v)}{P_f(u,v)+\sigma^2}$ where $\sigma^2$is variance from the noise $P_f(u, v)$is the signal power spectrum
Gabor	f is the frequency of sinusoidal function and θ is the orientation of Gabor function	$G(x,y) = \frac{f^2}{\pi\gamma\eta}\exp\left(-\frac{f^2}{\pi\gamma\eta}\right)\exp(j2\pi f x' + \phi)$ $x' = x\cos(\theta) + y\sin(\theta)$ $y' = y\cos(\theta) - x\sin(\theta)$
Haralick	Angular second moment	$\sum_i\sum_j p(i,j)^2$
	Contrast	$\sum_{n=0}^{N_g-1} n^2\left\{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)\right\}, \lvert i-j\rvert = n$
	Correlation	$\frac{\sum_i\sum_j (ij)p(i,j)-\mu_x\mu_y}{\sigma_x\sigma_y}$
	Sum of squares (variance)	$\sum_i\sum_j (i-\mu)^2 p(i,j)$
	Inverse different moment	$\sum_i\sum_j \frac{i}{1+(i-j)^p} p(i,j)$
	Sum average	$\sum_{i=2}^{2N_g} i p_{x+y}(i)$
	Sum varince	$\sum_{i=2}^{2N_g} (i-f_s)^2 p_{x+y}(i)$
	Sum entropy	$-\sum_{i=2}^{2N_g} p_{x+y}(i) log\left\{p_{x+y}(i)\right\}$
	Entropy	$-\sum_i\sum_j p(i,j)\log(p(i,j))$
	Difference variance	variance of p_{x-y}(i)
	Difference entropy	$-\sum_{i=0}^{N_g-1} p_{x-y}(i)\log\left\{p_{x+y}(i)\right\}$
	Info. measure of correlation 1	$\frac{HXY-HXY_1}{\max\{HX,HY\}}$
	Info. measure of correlation 2	$(1-\exp[-2.0(HXY2-HXY)])^{\frac{1}{2}}$ $HXY1 = -\sum_i\sum_j p(i,j)\log(p(i,j))$ where HX and HY are the entropies of p_x and p_y, $HXY1 = -\sum_i\sum_j p(i,j)\log\{p_x(i)(p)_y(j)\}$, and $HXY2 = -\sum_i\sum_j p_x(i)p_y(j)\log\{p_x(i)(p)_y(j)\}$
	Max. correlation coefficient	(the second largest eigenvalue of Q)$^{1/2}$ where $Q(i,j) = \sum_k \frac{p(i,k)p(j,k)}{p_x(i)p_y(k)}$
SFTA	$\Delta(x, y)$: SFTA feature vector	$\Delta(x,y) = \{1, \ 0, \text{ if}\exists\left(x', y'\right)\in N_4[x,y], \ Ib\left(x', y'\right) = 0\wedge Ib\left(x', y'\right) = 1, \text{otherwise}$ where $N_4[x, y]$ is the set of pixels that are 4-connected to (x, y) from the image. $\Delta(x, y)$ uses the value 1 if the pixel at position (x, y) in the binary image $Ib(x', y')$ that has the value 1 and having one neighboring pixel with the value 0. Otherwise, $\Delta(x, y)$ takes the value 0.

References

1. Ali GN, Chiang PJ, Mikkilineni AK, Chiu GT, Delp EJ, Allebach JP (2004) Application of principal components analysis and Gaussian mixture models to printer identification. In: Intl. Conference on digital printing technologies. Salt Lake City, pp. 301–305
2. Bekhti MA, Kobayashi Y (2016) Prediction of vibrations as a measure of terrain traversability in outdoor structured and natural environments, in: image and video technology, Vol. 9431 of the series lecture notes in

computer Science. Springer International Publishing, Auckland, pp 282–294. doi:10.1007/978-3-319-29451-3_23
3. Buchanan JDR et al (2005) Fingerprinting' documents and packaging. Nature 436:475. doi:10.1038/436475a
4. Bulan O, Mao J, Sharma G (2009) Geometric distortion signatures for printer identification International conference on acoustics, speech and signal processing (ICASSP). Taipei pp 1401-1404. doi: 10.1109/ICASSP.2009.4959855
5. Chiang PJ, Khanna N, Mikkilineni AK, Segovia MVO, Suh S, Allebach JP, Chiu GTC, Delp EJ (2009) Printer and scanner forensics: examining the security mechanisms for a unique interface. IEEE signal processing magazine. March, pp.72-83. doi: 10.1109/MSP.2008.931082
6. Choi JH, Lee HY, Lee HK (2013) Color laser printer forensic based on noisy feature and support vector machine classifier. Multimedia Tools Applications 67:363–382. doi:10.1007/s11042-011-0835-9
7. Chu PC, Cai BY, Tsoi YK, Yuen R, Leung KSY, Cheung NH (2013) Forensic analysis of laser printed ink by X-ray fluorescence and laser-excited plume fluorescence. Anal Chem 85(9):4311–4315. doi:10.1021/ac400378q
8. Chun-Lin L, (2010) A tutorial of the wavelet transforms. National Taiwan University http://disp.ee.ntu.edu.tw/tutorial/WaveletTutorial.pdf. Accessed 13 July 2016
9. Costa AF, Humpire-Mamani G, Traina AJM (2012) An efficient algorithm for fractal analysis of textures. SIBGRAPI Conference on Graphics, Patterns and Images, August, Ouro Preto. pp. 39–46. doi:10.1109/SIBGRAPI.2012.15
10. Cox IJ, Miller ML, Bloom JA, Fridrich J, Kalker T (2008) Digital watermarking and steganography, 2nd edn. Morgan Kaufmann Publishers, Amsterdam
11. Abramowitz M, Davidson MW, The concept of magnification. Olympus America, Inc. https://micro.magnet.fsu.edu/primer/anatomy/magnification.html. Accessed 2 Feb 2017
12. Ferreira A, Navarro LC, Pinheiro G, Santos JAD, Rocha A (2015) Laser printer attribution: exploring new features and beyond. Forensic Sci Int 247:105–125. doi:10.1016/j.forsciint.2014.11.030
13. Gonzales RC, Woods RE (2008) Digital Image Processing, 3rd edn. Prentice Hall, New Jersey
14. Haghighat M, Zonous S, Abdel-Mottaleb M (2015) CloudID: trustworthy cloud-based and cross-enterprise biometric identification. Expert Syst Appl 42(21):7905–7916. doi:10.1016/j.eswa.2015.06.025
15. Haralick RM, Shanmugam K, Dinstein I (1973) Textural features for image classification. IEEE Trans Syst, Man Cybern SMC-3(6):610–621
16. Herman B, Lemasters JJ (1993) Optical microscopy: emerging methods and applications. Academic Press, San Diego
17. Hewlett-Packard Company (2002) HP LaserJet 4200 and 4300 series printers http://www.nuigalway.ie/psy/sub/manuals/hp_lj4200.pdf. Accessed 21 June 2016
18. Hsu CW, Chang CC, Lin CJ (2003) A practical guide to support vector classification, Taipei: National Taiwan University. http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf. Accessed 18 July 2016
19. https://en.wikipedia.org/wiki/Digitizing. Accessed 2 Aug 2016
20. Jurič I, Ranđelović D, Karlović I, Tomić I (2014) Influence of the surface roughness of coated and uncoated papers on the digital print mottle. Journal of Graphic Engineering and Design 5(1): 17–23
21. Juuti M, Prykäri T, Alarousu E, Koivula H, Myllys M, Lähteelä A, Toivakka M, Timonen J, Myllylä R, Peiponen KE (2007) Detection of local specular gloss and surface roughness from black prints, in: colloids and surfaces a: physicochemical and engineering aspects. Elsevier 299(1-3):101–108. doi:10.1016/j.colsurfa.2006.11.039
22. Kawasaki M, Ishisaki M (2009) Investigation into the cause of print mottle in halftone dots of coated paper: effect of optical dot gain non-uniformity, vol.63. No.11, pp.1362–1373. http://www.tappi.org/content/06IPGA/5-4%20Kawasaki%20M%20Ishisaki.pdf. Accessed 27 June 2016
23. Kim DG, Lee HK (2014) Color laser printer identification using photographed halftone images, Proc. of EUSIPCO. September, IEEE, Lisbon, pp. 795–799
24. Kim KI, Jung K, Park SH, Kim HJ (2002) Support vector machines for texture classification. IEEE Trans Pattern Anal Mach Intell 24(11):1542–1550. doi:10.1109/TPAMI.2002.1046177
25. Kundur D, Lin CY, Macq B, H. Yu (2004) Scanning the issue: special issue on enabling security technologies for digital rights management, in Proceedings of the IEEE, pp. 879–882
26. Li Q, Zhang Z, Lu W, Yang J, Ma Y, Yao W (2016) From pixels to patches: a cloud classification method based on a bag of micro-structures. Atmospheric Measurement Techniques 9:753–764
27. Mäenpää T, Pietikäinen M (2004) Texture analysis with local binary patterns. In: Chen CH, Wang PSP (eds) Handbook of Pattern Recognition & Computer Vision, 3rd edn. World Scientific, Singapore, pp 115–118
28. Marcella AJ Jr, Guillossou F (2012) Cyber forensics: from data to digital evidence. John Willy & Sons, New Jersey

29. Mihlbachler MC, Beatty BL, Caldera-Siu A, Chan D (2012) Error rates in dental microwear analysis using light microscopy. Palaeontol Electron 15(12A):22
30. Mikkilineni AK, Chiang PJ, Ali GN, Chiu GTC, Allebach JP, Delp EJ (2005) Printer identification based on graylevel co-occurrence features for security and forensic applications. In Proceedings of the SPIE International Conference on Security, vol. 5681, pp. 430–440
31. Mikkilineni AK, Arslan O, Chiang PJ, Kumontoy RM, Allebach JP, Chiu GTC, Delp EJ (2005) Printer forensics using svm techniques in Proceedings of the IS&T's NIP21: International conference on digital printing technologies, vol. 21, Baltimore, October, pp. 223–226
32. Ojala T, Pietikäinen M, Mäenpää T (2002) Multiresolution gray-scale and rotation invariant texture classification with LBP, IEEE Trans. Pattern Analysis & Machine Intelligence 24(7): 971–987. doi:10.1109/TPAMI.2002.1017623
33. Oravec M, Gál L, Čeppan M (2015) Pre-processing of inkjet prints NIR spectral data for principal component analysis. Acta Chim Slov 8(2):191–196. doi:10.1515/acs-2015-0031
34. Osadchy M, Jacobs DW, Lindenbaum M (2007) Surface dependent representations for illumination insensitive image comparison. IEEE Trans Pattern Anal Mach Intell 29(1):98–111. doi:10.1109/TPAMI.2007.19
35. Pollard SB, Simske SJ, Adams GB (2010) Model based print signature profile extraction for forensic analysis of individual text glyphs. IEEE workshop on information forensics and security - WIFS'10, Seattle, December 12-15. http://www.hpl.hp.com/techreports/2010/HPL-2010-173.html. Accessed 6 Aug 2016
36. Pollard S, Simske S, Adams G (2013) Print Biometrics: Recovering forensic signatures from halftone images, Hewlett-Packard Development Company http://www.hpl.hp.com/techreports/2013/HPL-2013-1.pdf. Accessed 30 Jul 2016
37. Ryu SJ, Lee HY, Im DH, Choi JH, Lee HK (2010) Electrophotographic printer identification by halftone texture analysis. In: IEEE Intl. Conference on acoustics speech and signal processing (ICASSP). pp. 1846–1849. doi:10.1109/ICASSP.2010.5495377
38. Say OT, Sauli Z, Retnasamy V (2013) High density printing paper quality investigation, IEEE Regional Symposium on Micro and Nano electronics (RSM). Langkawi, pp. 273-277. doi:10.1109/RSM.2013.6706528
39. Schalkoff RJ (1989) Digital image processing and computer vision. John Wiley & Sons, Australia
40. Sharma G (2016) Image-based data interfaces revisited: barcodes and watermarks for the mobile and digital worlds. 8th International conference on communication systems and networks (COMSNETS). 5-10 January, 6 p. doi:10.1109/COMSNETS.2016.7440021
41. Sharma A, Subramanin L, Brewer E (2011) PaperSpeckle: microscopic fingerprinting of paper. Proceedings of the 18th ACM conference on Computer and communications security, Chicago, Illinois, USA — October 17–21, pp. 99–110 doi: 10.1145/2046707.2046721
42. Simske SJ, Adams G (2010) High-resolution glyph-inspection based security system, IEEE International Conference on Acoustics, Speech, and Signal Processing, 14–19 March, pp. 1794–1797. doi: 10.1109/ICASSP.2010.5495416
43. Su R, Pekarovicova A, Fleming PD, Bliznyuk V (2005) Physical Properties of LWC Papers and Gravure Ink Mileage https://www.researchgate.net/publication/251423637_Physical_Properties_of_LWC_Papers_and_Gravure_Ink_Mileage. Accessed 23 June 2016
44. Szynkowska MI, Czerski K, Paryjczak T, Parczewski A (2010) Ablative analysis of black and colored toners using LA-ICP-TOF-MS for the forensic discrimination of photocopy and printer toners. Survey and Interface Analysis 42:429–437. doi:10.1002/sia.3194
45. Tsai MJ, Liu J (2013) Digital forensics for printed source identification. In IEEE International Symposium on Circuits and Systems (ISCAS). May, pp. 2347–2350. doi: 10.1109/ISCAS.2013.6572349
46. Tsai MJ, Yin JS, Yuadi I, Liu J (2014) Digital forensics of printed source identification for Chinese characters. Multimedia Tools and Applications 73:2129–2155. doi:10.1007/s11042-013-1642-2
47. Tsai MJ, Hsu CL, Yin JS, Yuadi I (2015) Japanese character based printed source identification, IEEE International Symposium on Circuits and Systems (ISCAS). May, Lisbon. pp. 2800-2803. doi: 10.1109/ISCAS.2015.7169268
48. Vega LR, Rey H (2013) A rapid introduction to adaptive filtering. Springer-Verlag, Berlin Heidelberg
49. Voloshynovskiy S, Holotyak T, Bas P (2016) Physical object authentication: detection-theoretic comparison of natural and artificial randomness. IEEE International conference on acoustics, speech and signal processing, 20-25 March, pp. 2029-2033. doi: 10.1109/ICASSP.2016.7472033
50. Zhu B, Wu J, Kankanhalli MS (2003) Print signatures for document authentication. Proceedings of the 10th ACM conference on Computer and communications security. Washington D.C., USA October 27–30, 2003, pp. 145–154. doi: 10.1145/948109.948131

Min-Jen Tsai received the B.S. degree in electrical engineering from National Taiwan University in 1987, the M.S. degree in industrial engineering and operations research from University of California at Berkeley in 1991, the engineer and Ph.D. degrees in Electrical Engineering from University of California at Los Angeles in 1993 and 1996, respectively. He served as a second lieutenant in Taiwan army from 1987 to 1989. From 1996 to 1997, he was a senior researcher at America Online Inc. In 1997, he joined the institute of information management at the National Chiao Tung University in Taiwan and is currently a full professor. His research interests include multimedia system and applications, digital right management, digital watermarking and authentication, digital forensic, enterprise computing for electronic commerce applications. Dr. Tsai is a member of IEEE, ACM and Eta Kappa Nu.

Imam Yuadi is currently a Ph.D. student at the Institute of Information, National Chiao Tung University. He is also a lecturer and researcher in the Department of Information and Library Science, Airlangga University, Indonesia. Yuadi studied in the Library and Information Science from Padjadjaran University in 1999 and received his master degree in Information Technology Management from Sepuluh Nopember Institute of Technology at Surabaya in 2009. His research interest includes digital forensics, image recognition, computer vision and digital library.