A Survey on Techniques of Detecting Identity Documents Forgery

Alsadig Bashir Hassan
College of Computer Science and Technology
Sudan University of Science and Technology, SUST
Khartoum, Sudan
alsadigthree@gmail.com

Yahia A. Fadlalla
Lead Consultant/Researcher,
InfoSec Consulting,
Hamilton, Ontario, Canada.
trusted_software@usa.net

***Abstract*— Identity Document (ID) forgery is a process by which authorized identity document is modified and\or copied by unauthorized party or parties to be used illegally. The rapid development of personal computers, scanners, and color printers a long with their affordability raises up the risk of identity document forgery. They represent rich tools and techniques that highly support the creation of faked identities. This review paper investigates the current techniques for countering document forgery threats long with their achievements and limitations. These techniques are insufficient to mitigate ID forgery threats. New methods and techniques need to be developed to reduce or eliminate the risk of the Identity documents forgery.**

Keywords—IDcard, forgery, algorithim, watermarking, biometrics, steganography, printer Type.

I. INTRODUCTION

An Identity Document (ID) is used to identify a person or verify him or her aspects, e.g. name, age, address, ID number, etc. Some countries issue formal identity documents while others require identity verification using informal documents [1][2]. There are several types of IDs such as Passports, National Identification Cards, Residence Cards, Birth Certificates, Death Certificates, Driving Licenses, Military Identification, etc. [2][3][4]. The photo on the ID is used to connect a person a document.

ID fraud is built on the foundation of a fictitious identity, often created with a combination of real data and fabricated information. For example, the fraudster may (tomorrow) one's person Social Security number (SSN), combine it with another person's name, and use someone else's address to create a brand new identity. The perpetrator can then use this fraudulent identity to apply for credit, make major purchases, or a variety of other activities that give the identity a financial history.

Due to the advanced development of high quality computers, printers and scanners, which are comparatively of low cost, ID's forgery became a great problem nowadays. Fake photo IDs prevailed across the world, some example are illegal immigration, drug smugglers, human traffickers and terror attacks. Authors in [4] estimate that there will be bout 44% increase in frequent ID cards between 2014 to 2018, rising from $5 billion in annual losses to a projected $8 billion. Many researchers have discussed this threat and proposed different techniques for solving, resolving or avoiding it.

Therefore, in this survey paper, we reviewed the past and recent techniques that are used to discover ID forgery.

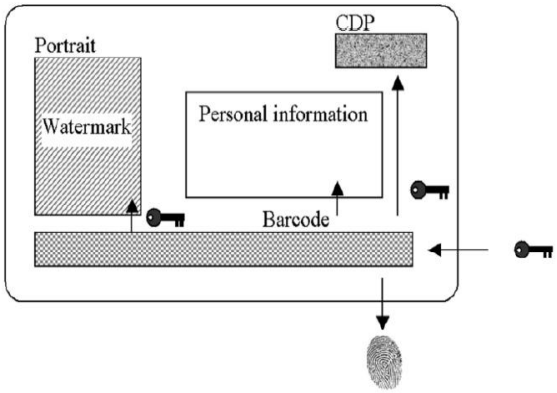

Fig. 1: Abstract of ID document with security features; CDP = Copy Detection Pattern [24]

The rest of this paper is organized as follows: Section II sheds lights on some techniques and methods that are commonly used in discovering fake IDs. Section III, presents a thorough investigation of the current techniques for countering document forgery threats and discusses their achievements and limitations. Section IV concludes.

II. FORGERY DETECTION TECHNIQUES

This section defines some techniques and methods that are commonly used in detecting IDs forgery.

(1) *Digital Watermarking:* It is a process that embeds data into a multimedia object to protect the one's ownership to the object. It can be classified into various categories: one that is, based on objects, e.g. , text, image, audio and

video. The other category is based on the human preception can be visible or invisible watermarking. A third category based on robustness, that is how the watermarking resists the attacks. robustness watermarking can be fragile, semi-fragile or robust [5], [6].

(2) *Biometrics:* They are unique physical or behavior characteristics that used to verify personal identity. Physical characteristics like the shape of the composition of the body such as fingerprint, DNA, face, hand, retina or odor. Behavioral characteristics are the behavior of the person, e.g., typing rhythm, gait, voice or gesture [10], [11].

(3) *Steganography:* It is a science of hiding information by embedding messages inside the text, image, audio, or video. Steganography's goal is to hide a message into an object [15], [16] .

III. RELATED WORKS

This section briefly reviews recent and past related work on techniques of detecting forgery in identity documents.

A fingerprint system is proposed in [19] to detect forged documents, which consist of fingerprint identification; reader; dataset; and physical ID artifact. The system collects fingerprints and matches them with existing ones that stored in a chip of the identity document. However, as the authors stated that their system faces some obstacles in case of wet, dry, or scratched fingerprint.

The authors in [20] discussed an approach for detecting forgery in Netherland passports and they stated its efficiency. This approach is based on using biometrics in forgery detection. However, the Netherland approach needs to be updated from time to time because face appearance of young people and old ones may change from time to time. Furthermore, changing the right fingerprint with left fingerprint and vice versa may affect the system performance.

An approach to figure out fraud photocopy documents using digital watermarking technique is introduced in [21]. The system is used to detect forgery in identity documents by using a unique personal code inserted inside the photo. This code added to the identity document without making any change or alternation in a photo. The system detects the forgery on the document by comparing the embedded data with the information on the identity document. However, digital watermarking has some weaknesses; (a) removal attack that aims to remove all watermarking; (b) cryptography attack that aims to alter watermarking, and protocol attack that aims to attack all watermarking applications.

In [22] a model of digital watermarking to authenticate the photo in Thai national ID card is proposed. The model has three steps: (1) embedding the watermarking in the owner photo; (2) print it on the ID card; and (3) extract the watermarking from the photo ID to determine if the information obtained is the same as the card ID code or not.

A new algorithm to remove the Printing and Scanning distortion (PS) of the watermarking on the ID cards is proposed in [23]. Commonly, PS occurs during the authentication process; it makes a noise on the ID. Therefore, the PS must be removed in order to complete the authentication process correctly. This indicates the important of developing such systems.

A system for forgery detection that is based on integration of four different techniques is presented in [24]. The proposed system consists of four components: watermarking, the 2D barcode, copy detection pattern (CDP), and biometrics. However, the system causes more overhead in term of computational process.

An approach to find out the fraud photocopied document by using Bounding Box (BB) is proposed in [25]. This approach concentrates on the part of document that has been altered by removing the original contents and writing above it or altering the contents through the cut and paste technique. BB uses Matlab tool to surround the characters and symbols regions to detect forgery on the documents by the height, orientation, and thickness on the document, if there is a difference in the suspected part of the document, then the system detects a forgery. However, as this approach may take too much time during the comparison process.

One of the common methods in detecting the passport forgery is based on the Machine Readable Zone (MRZ) is presented in [26]. MRZ is very essential in passport recognition, which is placed on the bottom of the passport. If the data in the passport compared with MRZ code is not identical, that means the passport is fabricated.

In [27] a technique for detecting counterfeited document, Scan-Edit and Print (SEP), is presented. SEP consists of two components: (i) copy and paste detector which can detect copy and paste forgery, by which a group of characters are copied and pasted in different location on the documents; and (ii) the imitation detector that detects an imitation forgery where the counterfeiter adds or alters the information by imitation of the font properties. This technique uses the intrinsic document feature to detect the forgery in documents, such as character font properties, the shape of the character and the character or words alignments. However, this system would be more efficient if a technique to reduce the print and scan noise is added.

In [28], researchers provided a solution for detecting fabricated text-documents, which is called Copy-Move Forgery Detection (CMFD). The copy-move is to copy a part of text-document and reinsert it again in another location in the same text-document. This aims to hide undesirable contents or duplicating special part, e.g., to fabricate names, dates or values. CMDF consists of two parts: (1) Optical Character Recognition (OCR), focuses on detecting font forgery by measuring their weight, size, style and roughness. (2) Copy Move (CM), which is focusing on the background of the text-document. However, CMDF would be more efficient if utilizes a Conditional Random technique.

The Conditional Random Field (CRF) system is proposed in [29]. CRF is a tool that describes the correlation between fonts, styles, and sizes of the characters. The system decides which font type the character belong to by comparing the character font features to a predefined database, and classify them to know whether the character is genuine or fake. The system detects three types of errors: (a) copy/paste error that is space between a pair of characters different from the rest of the characters on the document; (b) imitation error which is the words that appear in two different typestyles; and (c) copy/paste and imitation error which is the two above errors found in one document.

Table 1. Some experimental results to compare the performance of surveyed work

Year	Author\s	Title	Technique\s	Results
2007	Young Bin Kwon and et al.	Recognition based Verification for the Machine Readable Travel Documents.	Machine Readable Zone (MRZ)	The recognition of this system was 99.8% of detecting forgery documents.
2008	Abbas Chedad and et al.	Combating digital document forgery using new secure information hiding algorithm	Steganography	Efficiency of 75.0% was achieved by this system in detecting forgery identity documents.
2009	Christian Schulze and et al	Using DCT features for printing technique and copy detection.	Discrete Cosine Transformation and Machine Learning	The accuracy was achieved by this system was 90.0% of detecting forgery documents.
2013	Suman V Patgar and et al.	An Unsupervised Intelligent System to Detect Fabrication in Photocopy Document Using Geometric Moments and Gray Level Co-Occurrence Matrix.	Geometric Moments and Gray Level Co-Occurrence Matrix	Efficiency of 94.59% was achieved by the approach in detecting forgery photocopied documents.
2014	Sara Elkasrawi and et al.	Printer Identification using Supervised Learning for Document Forgery Detection	An approach to detect the source of the printers.	Accuracy of whole dataset was 76.75%.
2014	Clarisse MANDRIDAKE and et al.	Towards Fully Automatic ID Document frauds detection	Special ink for detect forgery on background	The accuracy was achieved by this system was 68.0% of detecting forgery documents.
2014	Suman V Patgar and et al.	An unsupervised intelligent system to detect fabrication in photocopy document using variations in Bounding Box features	Bounding Box tool	Efficiency of 85.7% was achieved by this system in detecting forgery photocopied documents.
2016	Svetlana Abramova and et al.	Detecting Copy–Move Forgeries in Scanned Text Documents	Copy–Move Forgery Detection (CMFD)	The accuracy was achieved by this system was 87.0% of detecting forgery documents.

The authors in [30] utilize a method to detect fabricated photocopied documents by Geometric Moments and Gray

Level Co-occurrence Matrix Features. They focus on detecting fabrication that occurs in the document in which some contents of the original contents have been manipulated or overwritten. The system can detect the forgery in a document by its consistent intensity; if it is smooth and has strong edge contour, it indicates a non-fabricated photocopied text; if it is rough or has a weak edge contour, it indicates a fabricated photocopied text. The system in [30] is also use to enhance the documents by making it free of noise and dirt.

In [31] an approach called Support Vector Machine (SVM) has been trained to distinguish the laser printout from the inkjet printout. The proposed approach is used to detect the forgery in documents by focusing on the edge areas of the letters. The system determines which type of printer has been used to print a document, if the printed letters have sharper edges, this indicates a laser printer while non-sharper edges indicates inkjet printer.

The authors in [32] introduced a framework to detect the forgery on the documents by focusing on the static part of the printed document. Most of the printed documents have a

static part (header and footer) and a non-static part (the actual content of the document). This framework compares the static parts together to determine if the printed document fakes or not. However, by adding the non-static part to the authentication process, the system would be more accurate.

In [33], an approach to distinguish the laser printed documents from inkjet printed documents is proposed. It unsupervised anomaly detection that does not need prior training. The system used large dataset of inkjet and laser printed documents. However, by adding like Conditional Random Field (CRF) it could be more efficient.

An approach to detect the source of the printers is presented in [34] it relies on the noise produced by the printer, regardless of the content and the size of the document. The proposed approach differentiates the inkjet printer form other printers.

In [35] a method to detect forgery on scanned documents based on hiding techniques is presented. Steganography is one of the data hiding techniques; it is the science of invisible imbedding of data in a digital medium. However, Steganography's systems require a lot of overhead to hide relatively few bits of data.

A system that uses to classify the font type Condition Random Field (CRF), proposed in [36]. The system aims to identify the typeface, weight, slope, and size of the fonts without knowing the content of the text. The system ignores punctuations and automatically predicts the font type by comparing the typeface, slope, weight, and the size of the fonts. However, the system could be more efficient for detection process by adding another technique, like printer signature.

In [37], the authors offered an algorithm to detect the documents forgery by compute the document alignment (left, right, justified, and centered), In addition, the algorithm also checks the skew and analyses the text-lines to enrich the algorithm capability to detect forged document. However, this approach deals only with a pure text document. It fails to detect image forgery.

A system to detect the printer type by using Discrete Cosine Transformation (DCT) and Machine Learning Technique is introduced in [38]. The system helps to differentiate between the inkjet and laser printers based upon two things: (1) The characters edges that produced by the laser printer are sharper, while the inkjet printer produces a tendency and blurring character edges; (2) The edges roughness and degradation for the inkjet printed documents have high degree of edges roughness while the laser-printed documents have a tiny edge roughness.

An approach to detect the forged documents by focusing on the background area and photo area is proposed in [39]. This approach detects special kind of ink that was used during the printing process. On the other hand, the system detects forgery on printed photo by identifying the printer used to print that photo. Inkjet printers indicate forged documents and laser printers indicate original documents.

Authors in [40] provided a solution to detect forged documents by using an image texture analysis to classify the printer used. The system uses two methods to identify the printer type; the first method identifies a printer according to the basic characteristics in a printed document; such as the roughness and the sharpness of the characters. This method is known as intrinsic method. The second method embeds the external signature in the printed documents; e.g., a printer's serial number, date of printing. This method referred to as extrinsic method.

IV. SUMMARY

The Detection of fake documents is considered a critical issue in the information security field. The fabrication of identity documents has received a highly attention by the research community due to its direct connection with dangerous crimes – illegal immigration, financial fraud, human and drug smuggling, terrorism, etc. The rapid development of network technologies assists hackers to easily get high quality fake IDs through different Web sites, and this opened doors widely for fake IDs markets [2] [3]. Some reports show that 60% of fake documents can be detected through detection machines or methods while 80% can be detected by human experts [39]. This indicates that many fake documents have not yet been detected. Therefore, much work should be done to enhance and develop efficient forgery detection methods.

The methods and techniques reviewed and discussed have drawbacks, therefore, more work in this field is highly encourage to mitigate the risk behind these threats, more ID detections techniques must be sought and made available.

The employment of cryptography (e.g., watermarking) needs to be encouraged in detecting ID forgery or making existing ID-making techniques more secure.

REFERENCES

[1] National Document Fraud Unit (UK Home Office), Guidance on examining identity documents, London, Britain, 2016.

[2] National Document Fraud Unit (UK Home Office), Guidance on examining identity documents, London, Britain, 2015.

[3] Cifas (Fraud Prevention), ID Documents Report, London, Britain, 2014.

[4] Equifax and EFX, "The New Reality Synthetic ID Fraud," Atlanta, Georgia, 2015.

[5] Cox, Ingemar J., et al. *Digital watermarking*. Vol. 1558607145. San Francisco: Morgan Kaufmann, 2002.

[6] Channapragada, Rama Seshagiri Rao, Anil Srimanth Mantha, and Munaga VNK Prasad. "Study of contemporary digital watermarking techniques." *International Journal of Computer Science Issues(IJCSI)* 9.6 (2012).

[7] Akter, Afroja, and Muhammad Ahsan Ullah. "Digital watermarking with a new algorithm." *International Journal of Research in Engineering and Technology* 3 (2014).

[8] Nasereddin, Hebah HO. "Digital watermarking a technology overview." *International Journal of Research and Reviews in Applied Sciences* 6.1 (2011): 89-93.

[9] Singh, O. P., et al. "Study of Watermarking Techniques Used in Digital Image." *International Journal of Scientific and Research Publications* 2.10 (2012).

[10] Uddin, Mohammed Nasir, et al. "A Survey of Biometrics Security System." *IJCSNS* 11.10 (2011): 16.

[11] Andronikou, Vassiliki, Dionysios S. Demetis, and Theodora Varvarigou. "Biometric implementations and the implications for security and privacy." *Journal of the Future of Identity in the Information Society* 1.1 (2005): 20-35.

[12] UK Biometric Working Group. *Biometric security concerns*. Technical Report, CESG, September 2003, http://www. cesg. gov. uk/site/ast/biometrics/media/BiometricSecurityConcerns. pdf, 2003.

[13] Matyáš, Václav, and Zdeněk Říha. "Biometric authentication—security and usability." *Advanced Communications and Multimedia Security*. Springer US, 2002. 227-239.

[14] Orsag, Filip, and Martin Drahanský. "Biometric Security Systems: Fingerprint and Speech Technology." *IICAI*. 2003.

[15] Bhattacharyya, Souvik, Indradip Banerjee, and Gautam Sanyal. "A survey of steganography and steganalysis technique in image, text, audio and video as cover carrier." *Journal of global research in computer science* 2.4 (2011): 1-16.

[16] Al-Othmani, Abdulaleem Z., Azizah Abdul Manaf, and Akram M. Zeki. "A survey on steganography techniques in real time audio signals and evaluation." *International Journal of Computer Science Issues (IJCSI)* 9.1 (2012).

[17] Cheddad, Abbas, et al. "Digital image steganography: Survey and analysis of current methods." *Signal processing* 90.3 (2010): 727-752.

[18] Li, Bin, et al. "A survey on image steganography and steganalysis." *Journal of Information Hiding and Multimedia Signal Processing* 2.2 (2011): 142-172.

[19] Yang, Chunlin. "Fingerprint biometrics for ID document verification." 2014 9th IEEE Conference on Industrial Electronics and Applications. IEEE, 2014.

[20] Schouten, Ben, and Bart Jacobs. "Biometrics and their use in e-passports."Image and Vision Computing 27.3 (2009): 305-312.

[21] Perry, Burt, Scott Carr, and Phil Patterson. "Digital watermarks as a security feature for identity documents." PROC SPIE INT SOC OPT ENG. Vol. 3973. 2000.

[22] Thongkor, Kharittha, and Thumrongrat Amornraksa. "Digital image watermarking for photo authentication in Thai national ID card." Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2012 9th International Conference on. IEEE, 2012.

[23] Ibrahim, Subariah, Masoud Afrakhteh, and Mazleena Salleh. "Adaptive watermarking for printed document authentication." Computer Sciences and Convergence Information Technology (ICCIT), 2010 5th International Conference on. IEEE, 2010.

[24] Picard, Justin, Claus Vielhauer, and Niels Thorwirth. "Towards fraud-proof id documents using multiple data hiding technologies and biometrics."Electronic Imaging 2004. International Society for Optics and Photonics, 2004.

[25] Patgar, Suman V., K. Rani, and T. Vasudev. "An unsupervised intelligent system to detect fabrication in photocopy document using variations in Bounding Box features." Contemporary Computing and Informatics (IC3I), 2014 International Conference on. IEEE, 2014.

[26] Kwon, Young-bin, and Jeong-hoon Kim. "Recognition based Verification for the Machine Readable Travel Documents." International Workshop on Graphics Recognition (GREC 2007), Curitiba, Brazil. 2007.

[27] Bertrand, Romain, et al. "A system based on intrinsic features for fraudulent document detection." 2013 12th International Conference on Document Analysis and Recognition. IEEE, 2013.

[28] Abramova, Svetlana. "Detecting Copy–Move Forgeries in Scanned Text Documents." Electronic Imaging 2016.8 (2016): 1-9.

[29] Bertrand, Romain, et al. "A Conditional Random Field model for font forgery detection." Document Analysis and Recognition (ICDAR), 2015 13th International Conference on. IEEE, 2015.

[30] Patgar, Suman V., and T. Vasudev. "An Unsupervised Intelligent System to Detect Fabrication in Photocopy Document Using Geometric Moments and Gray Level Co-Occurrence Matrix." International Journal of Computer Applications 74.12 (2013).

[31] Lampert, Christoph H., Lin Mei, and Thomas M. Breuel. "Printing technique classification for document counterfeit detection." 2006 International Conference on Computational Intelligence and Security. Vol. 1. IEEE, 2006.

[32] Ahmed, Amr Gamal Hamed, and Faisal Shafait. "Forgery detection based on intrinsic document contents." Document Analysis Systems (DAS), 2014 11th IAPR International Workshop on. IEEE, 2014.

[33] Gebhardt, Johann, et al. "Document authentication using printing technique features and unsupervised anomaly detection." 2013 12th International Conference on Document Analysis and Recognition. IEEE, 2013.

[34] Elkasrawi, Sara, and Faisal Shafait. "Printer identification using supervised learning for document forgery detection." Document Analysis Systems (DAS), 2014 11th IAPR International Workshop on. IEEE, 2014.

[35] Cheddad, Abbas, et al. "Combating digital document forgery using new secure information hiding algorithm." Digital Information Management, 2008. ICDIM 2008. Third International Conference on. IEEE, 2008.

[36] Satkhozhina, Aziza, Ildus Ahmadullin, and Jan P. Allebach. "Optical font recognition using conditional random field." Proceedings of the 2013 ACM symposium on Document engineering. ACM, 2013.

[37] van Beusekom, Joost, Faisal Shafait, and Thomas M. Breuel. "Text-line examination for document forgery detection." International Journal on Document Analysis and Recognition (IJDAR) 16.2 (2013): 189-207.

[38] Schulze, Christian, et al. "Using DCT features for printing technique and copy detection." IFIP International Conference on Digital Forensics. Springer Berlin Heidelberg, 2009.

[39] MANDRIDAKE, Clarisse, et al. "Towards Fully Automatic ID Document frauds detection."

[40]] Mikkilineni, Aravind K., et al. "Printer identification based on texture features." NIP & Digital Fabrication Conference. Vol. 2004. No. 1. Society for Imaging Science and Technology, 2004.