

A SYSTEM FOR DETECTION OF FABRICATION IN PHOTOCOPY DOCUMENT

Suman V Patgar, Vasudev T and Murali S

Maharaja Research Foundation,
Maharaja Institute of Technology Mysore,
Belawadi, S.R Patna, Mandya, India, 571438
sumanpatgar@gmail.com

ABSTRACT

Photocopy documents are very common in our normal life. In India, people are permitted to carry and produce photocopied documents frequently, to avoid damages or losing the original documents. But this provision is misused for temporary benefits by fabricating fake photocopied documents. When a photocopied document is produced, it may be required to check for its originality. An attempt is made in this direction to detect such fabricated photocopied documents. This paper proposes an unsupervised system to detect fabrication in photocopied document. The work in this paper mainly focuses on detection of fabrication in photocopied documents in which some contents are manipulated by new contents above it using different ways. Testing is performed with a different set of collected testing samples resulted in an average detection rate of 70.14%.

KEYWORDS

Document, Photocopy, Fabrication, Unsupervised system

1. INTRODUCTION

Many authorities in India trust and consider the photocopied documents submitted by citizens as proof and accept the same as genuine. Few such applications like to open bank account, applying for gas connection, requesting for mobile sim card, the concerned authorities insist photocopy documents like voter id, driving license, ration card, pan card and passport as proof of address, age, photo id etc to be submitted along with the application form. Certain class of people could exploit the trust of such authorities, and indulge in forging/ tampering/ fabricating photocopy document for short / long term benefits unlawfully. Fabricated photocopy is generated normally by making required changes intelligently in the photocopy obtained from an original document and then taking the recursive photocopy [1] from the modified photocopy. It is learned that in majority cases, fabrications are made by replacing a different photograph in place of photograph of an authenticated person; replacing contents in variable regions [2] through cut-and-paste technique from one or more documents; overlaying new content above actual content; adding new content into existing content; removing some content from existing; changing content by

overwriting; intellectually changing character in contents. It is quite evident from the applications listed above, the fabrication could be mainly made in the variable regions of documents.

The fabricated photocopy documents are generated to gain some short term or long term benefits unlawfully. This poses a serious threat to the system and the economics of a nation. In general, such frauds are noticed in the application areas where photocopy documents are just enough. These types of systems trusting photocopied document raise an alarm to have an expert system [3] that efficiently supports in detecting a forged photocopy document. The need of such requirement to the society has motivated us to take up research through investigating different approaches to detect fabrication in photocopy document.

2. RELATED WORKS

In literature no significant effort is noticed towards research on fabricated photocopy documents except for the work of detection of fabrication in photocopy document using GLCM features [4]. In this work, attempt is made to detect fabrication of photocopy documents in which text in variable region is fabricated by putting new contents in many ways. Many research attempts are carried out on original documents instead on photocopied documents, like signature verification, detection of forged signature [5], handwriting forgery [6], printed data forgery [7], and finding authenticity of printed security documents [8]. Literature survey in this direction reveals that the above research attempts have been made in the following issues: Discriminating duplicate cheques from genuine ones [8] using Non-linear kernel function; Detecting counterfeit or manipulation of printed document [7] and this work is extended to classify laser and inkjet printouts; Recognition and verification of currency notes of different countries [9] using society of neural networks along with a small work addressing on forged currencies; Identification of forged handwriting [6] using wrinkles as a feature is attempted along with comparison of genuine handwriting. The domain of research is in its early stage and there is no standard data set available for experimentation. The fabricated samples are obtained through writing new contents over an actual contents, smeared whitener, cut and paste and adding new contents above it.

The organization of the paper is as follows. In Section 3, the proposed model is explained with the support of a block diagram along with a brief introduction. The experimental results and a qualitative comparative analysis on the state of the art techniques are discussed in Section 4. The work is concluded in Section 5.

3. PROPOSED MODEL

Figure 1 shows the flow sequence of stages in the fabrication detection process. The system takes the photocopy document as input. The photocopy documents are the one which are subjected for detection of fabrication in textual region. The input photocopy document is scanned at 300 dpi to obtain bitmap image. The image is preprocessed to minimize the noise using median filter. The next stage considers the image and segment variable regions from the document image for detection of fabrication in the segmented area [12]. The segmented part is used to detect fabrication in the variable region [4,10,13].

As the area/ region of fabrication in the document is relatively smaller, compared to entire document it is quite hard to take decision whether the document is fabricated or non-fabricated. Hence it is favorable to employ all the three methods to check for the fabrication [4,10,13]. The

content in the variable region is decided as fabricated when one or more of the approaches lead to the classification as fabricated class. The content in the variable region is decided as non fabricated when all the three approaches confirm the non fabrication. The rejection cases in the method are considered as conflict and suggested for manual verification.

It is quite clear from the Figure 1, estimation of recursive order number of the photocopy is an input to the validation stage [1]. When the detection stage indicates the photocopy is not fabricated and the recursive order number is 1 then it is confirmed that the photocopy is not fabricated and the photocopy is considered as genuine photocopy. When the detection stage indicates fabrication in photocopy and the recursive order number is 2 or above then it can be confirmed that the photocopy is fabricated. In such case, the photocopy submitted can be rejected. Remaining combinations are the conflict cases and forwarded for manual verification with the original document at a later point in time. The decision rule for validation is illustrated in Figure 2 in the form of decision tree.

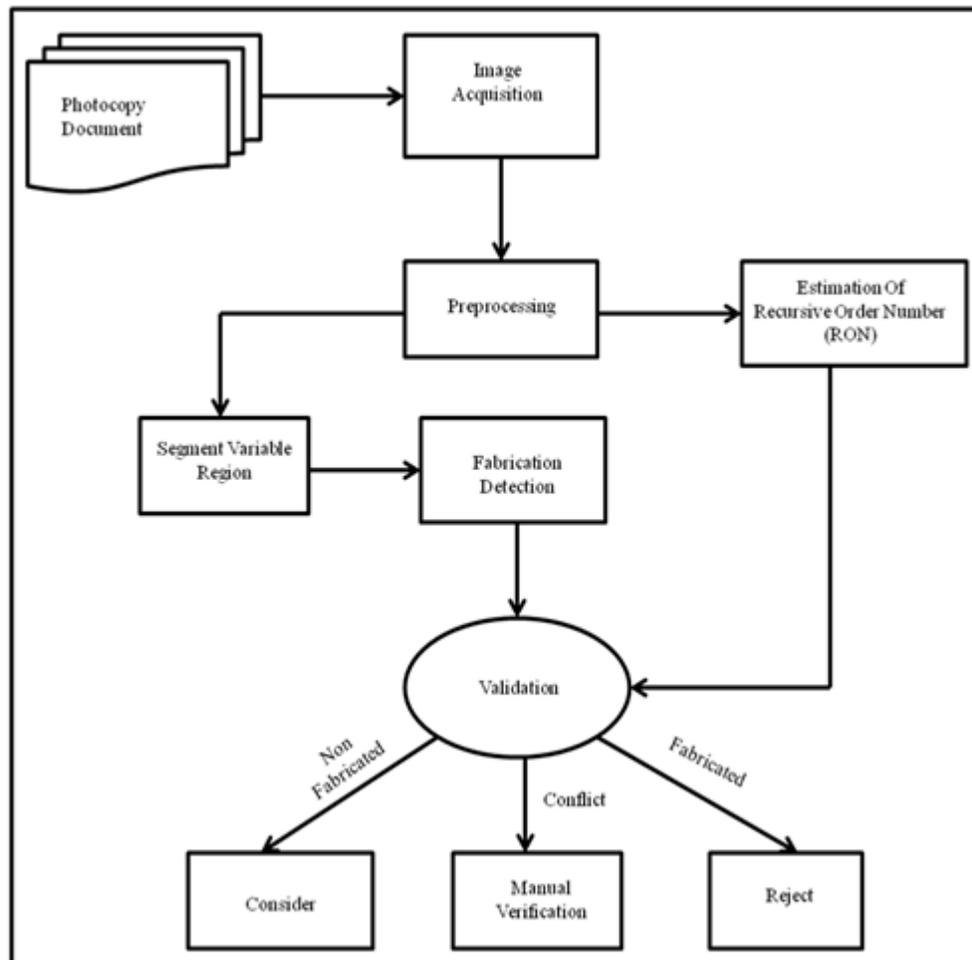


Figure 1: Flow sequence of stages in Detection of Fabrication

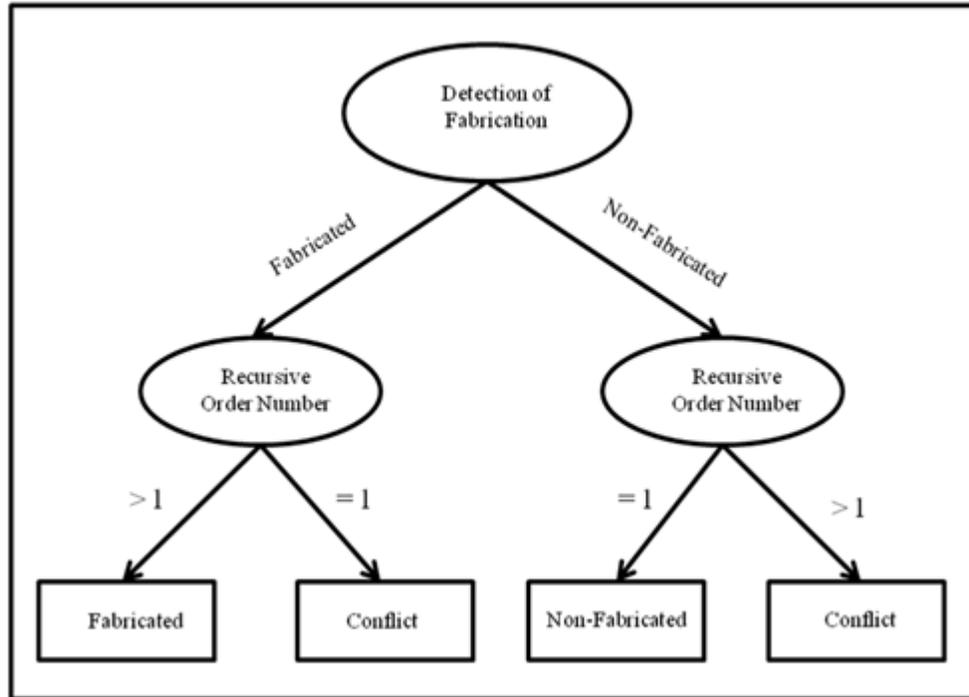


Figure 2: Decision Tree for Validation

4. EXPERIMENTAL RESULT

The proposed model is tested through considering 134 numbers of samples consisting of fabricated and non fabricated photocopy documents generated synthetically. The samples include DDs, certificates, birth / death certificates, cheques, attendance certificates, participation certificate and degree certificates obtained at different recursive orders. Samples are selected on which variable regions are underlined. Table (1) through (3) show results obtained at each stage.

Table 1 illustrates the results of 1st stage. This step shows correct segmentation efficiency as 82.83%. Table 1 indicates that 23 out of 134 samples are incorrect segmentation. The system is very much sensitive for improper segmentation. Every segmented region is forwarded to the detection of fabrication stage.

Table 1: Segmentation of Variable Region Result

No. of Samples	Correct Segmented	Incorrect Segmented	Segmentation Efficiency
134	111	23	82.83%

Table 2 shows the results of the fabrication detection stage. The segmented regions of the previous stage are the input for this stage.

Table 2: Detection of Fabrication Result

No. of samples	No. of correct classified	No. of Misclassified	Classification Efficiency
134	98(73.13%)	36(26.87%)	73.13%

The validation stage attempts to confirm the decision of fabrication detection stage and re-tunes the misclassification through classifying them as to correct class or conflict class. The validation stage considers fabrication detection output and recursive order number of the photocopy as input for validation. The Table 3 shows the result of detection of fabrication after validation.

Table 3: Validation Result

No. of samples	No. of correct classified	No. of Conflicts	Validation Efficiency
134	94(70.14%)	40(29.86%)	70.14%

The system is very much sensitive for improper segmentation. Because of the incorrect segmentation in input a considerable amount of drop in efficiency is noticed in this stage.

5. CONCLUSION

A complete integrated system has been proposed in this paper to detect fabrication in photocopy documents. The system takes the photocopy document as input and goes through various stages for classification of photocopy document. An overall performance of the system is about 70.14%. The reduction of efficiency in the system is due to efficiency drop happening at each stage in the system. The domain of the problem is new and challenging to come out with more generic solution especially in the stage of segmentation of variable region in the document. Scope to explore more efficient models to investigate fabrication in variable region is still open for research. The main limitation in the system is failures in estimating lower order recursive order number. More accurate models are required to detect lower recursive order number. More perfection is required in segmentation of variable region from documents. The automatic detection model has certain limitations like failure in detecting fabrication in the minute change in first order photocopy document and removing or adding few leading or trailing characters. There is scope to improve the efficiency of detection of fabrication by strengthening the model through improving segmentation of variable region in photocopy document.

REFERENCES

- [1] Suman Patgar, Vasudev T, Estimation of Recursive Order Number of a Photocopied Document Through Entropy From Gray Level Co-occurrence Matrix, 2013.
- [2] Vasudev T, Automatic Data Extraction from Pre-Printed Input Data Forms: Some New Approaches, PhD thesis supervised by Dr. G. Hemanthakumar, University of Mysore, India, 2007.

- [3] Rich Kevin Knight, Artificial Intelligence, 2nd Edition, McGraw-Hill Higher Education.
- [4] Suman Patgar, Vasudev T, 2013 An Unsupervised Intelligent System to Detect Fabrication in Photocopy Document Using Geometric Moments and Gray Level Co-Occurrence Matrix, IJCA(0975-8887) volume- 74/NO. 12, July 2013.
- [5] Madasu Hanmandlu, Mohd. Hafizuddin Mohd. Yusof, Vamsi Krishna Madasu off-line signature verification and forgery detection using fuzzy modeling Pattern Recognition Vol. 38, pp 341-356, 2005.
- [6] Cha, S.-H., & Tapert, C. C., Automatic Detection of Handwriting forgery, Proc. 8thInt.Workshop Frontiers Handwriting Recognition(IWFHR-8), Niagara, Canada, pp 264-267, 2002.
- [7] Christoph H Lampert, Lin Mei, Thomas M Breuel Printing Technique Classification for Document Counterfeit Detection Computational Intelligence and Security, International Conference, Vol. 1, pp 639-644, 2006.
- [8] Utpal Garian, Biswajith Halder, On Automatic Authenticity Verification of Printed Security Documents, IEEE Computer Society Sixth Indian Conference on Computer vision, Graphics & Image Processing, pp 706-713, 2008.
- [9] Angelo Frosini, Marco Gori, Paolo Priami, A Neural Network-Based Model For paper Currency Recognition and Verification IEEE Transactions on Neural Networks, Vol. 7, No. 6, Nov 1996.
- [10] Suman V Patgar, Y H Sharath Kumar, Vasudev T, Detection of Fabrication in Photocopy Document Using Texture Features through K-Means Clustering, Signal & Image Processing : Signal & Image Processing: An International Journal (SIPIJ), Vol.5, No.4, pp: 29-36, August 2014.
- [11] Suman V Patgar, Rani K, Vasudev T, A Voting based Approach to Detect Recursive Order Number of Photocopy Documents using Probability Distributions, Signal & Image Processing : An International Journal(SIPIJ), Vol.5, No.4, pp: 37-44, August 2014.
- [12] Suman V Patgar, Y H Sharath Kumar, Vasudev T, Segmentation of Handwritten Text from Underlined Variable Regions in Documents, International Journal of Information Processing(IJIP), Volume 8, issue 3, pp:23-30, 2014
- [13] Suman V Patgar, Rani K, Vasudev T, An unsupervised intelligent system to detect fabrication in photocopy document using Variations in Bounding Box Features, accepted in International conference on Contemporary Computing and Informatics (IC3I), 2014, Mysore, India

AUTHORS

Dr. Vasudev T is Professor, in the Department of Computer Applications, Maharaja Institute of Technology, Mysore. He obtained his Bachelor of Science and post graduate diploma in computer programming with two Masters Degrees one in Computer Applications and other one is Computer Science and Technology. He was awarded Ph.D. in Computer Science from University of Mysore. He is having 30 years of experience in academics and his area of research is Digital Image Processing specifically document image processing and authored about 50 research papers.



Dr. Mural S has obtained his Bachelors, Masters and PhD degrees from University of Mysore. Presently he is working as Professor in the Department Computer Science and Engineering, Maharaja Institute of Technology, Mysore, Karnataka, India. He has a total experience of 30+ years experience in teaching and about 20 years of experience in research. His area of research is Computer Vision and authored nearly 200 research papers and articles. He has served at different capacities as member and chairman in various academic bodies of Universities, UGC and AICTE. He is the Founder President of Maharaja Educational Trust, Mysore having around 10 different academic institutions.



Dr. Suman V Patgar, obtained her Bachelor of Engineering from Kuvempu University in 1998. She received her Masters degree in Computer Science and Engineering from VTU Belgaum in 2004. She was awarded Ph.D. in Computer Science from University of Mysore in 2015. She is having 10 years of experience in academics and her area of research is Digital Image Processing specifically document image processing and authored about 10 research papers.

