

# Automatic Fogery Detection on Printed Invoices: Results and Conclusions

Joost van Beusekom, Armin Stahl,  
German Research Center for Artificial Intelligence (DFKI)

The wide-spread use of high-quality scanning and printing devices, as well as the increasing degree of automation of the incoming mail processing, has opened many opportunities for the fraudulent use of forged documents. Due to the tremendous amount of content being processed, manual first-line inspection is not feasible anymore. Therefore, methods using intrinsic document features have been developed to authenticate documents coming from the same source.

We present the results and conclusions of a large-scale study on over 100,000 real-world invoices in the health sector, where we investigated a variety of innovative authentication techniques:

- **Text-line analysis** [1]: the idea of the text-line analysis is to measure text-line rotation angle and alignment to identify text-lines showing abnormal variations (which might have been manipulated). Despite the good results on synthetic data, it was noticed that on the real-world data the tested method is of little use.
- **Printing technique recognition and copy detection** [2]: in this approach, low-level image features are used to classify a print-out according to its printing technique. Different printing techniques inside one document source are considered suspicious. For this method, it could be shown that the results depend much more on non-printing technique related variations as expected. However, a new, promising approach could be identified that is currently under development.
- **Statistical layout models** [3]: using document alignment methods, distortions between documents from the same source are being measured. These distortions can come from a scanning or printing process and can thus be used to identify forged documents. It could be shown that this method gives reasonable results, but also that the nature of the underlying distortions in practice is not yet fully understood.
- **Counterfeit protection system codes** [4]: this approach employs the tiny yellow dots printed by color laser printer and color copiers for identification of the document source. This method proved to work well. One drawback, however, is the relatively rare appearance of the dots in the real-world dataset.

## References

- [1] J. van Beusekom, F. Shafait, T. M. Breuel, Text-line examination for document forgery detection, *Int. Jour. on Document Analysis and Recognition* (2012) 1–19.
- [2] C. Schulze, M. Schreyer, A. Stahl, T. M. Breuel, Evaluation of graylevel-features for printing technique classification in high-throughput document management systems, in: *Proc. of the 2nd Int. Workshop on Computational Forensics*, Vol. 5158 of *Lecture Notes in Computer Science*, Washington, DC, USA, 2008, pp. 35–46.
- [3] J. van Beusekom, F. Shafait, Distortion measurement for automatic document verification, in: *Proc. of the 11th Int. Conf. on Document Analysis and Recognition*, Beijing, China, 2011.
- [4] J. van Beusekom, M. Schreyer, T. M. Breuel, Automatic counterfeit protection system code classification, in: *Proc. of SPIE Media Forensics and Security XII*, San Jose, CA, USA, 2010.