University of Mosul

College of Computer Science and Mathematics

Detecting Forgery Documents Using Spectroscopy: A Network-Based Model

Alaa Amjad Salih

M.Sc. Thesis

Computer Science

Supervised by

Dr. Basim Mohammad Mahmood

Assistant Professor

1443 A.H **2021 A.D**

Abstract

The current technological era is witnessing a great revolution in the development of documents-processing applications. They are used for a variety of purposes such as producing and editing documents. Furthermore, the availability of printers (e.g., laser and inkjet printers) makes it easy for individuals to use them without restrictions. These printers and applications can be illegally exploited in manipulating or imitating original documents by forgers.

This thesis suggests a novel unsupervised forgery detection framework that is able to distinguish whether a document is forged or original based on spectroscopy of documents ink. The spectrums of the tested documents inks (original and questioned) are obtained using the Laser-Induced Breakdown Spectroscopy (LIBS) technology.

A correlation matrix of the spectrums is calculated for both the original and questioned documents together. Thereafter, the correlation matrix is transformed into an adjacency matrix aiming at converting it into a weighted network under the concept of graph theory. In the generated weighted network, the LIBS spectrums represent the vertices (V) for the network, and the correlations among the spectrums represent the weights (w) of the edges (E) among the corresponding vertices. After generating a weighted network for the original and questioned documents, a Louvain clustering algorithm is applied. Hence, the detection rate is measured based on the number of the generated clusters in the network. The proposed approach is tested under a variety of documents (e.g., official transcripts, official letters) and different types of printers (e.g., inkjet, laser, and photocopiers) as well as different kinds of papers. The findings show that the proposed approach provides 95.6% of accuracy in distinguishing printed samples from different printers using the Louvain clustering algorithms. Furthermore the proposed approach provides knowing the ink type used in print the questioned document and the printer's model that was used to print it. The obtained results are also benchmarked with other algorithms in the literature. Furthermore, the results of

the proposed approach can be visually interpreted, which is more comfortable for investigators who are not professionals in computer systems. Finally, the proposed approach is considered simple and does not need complex computations compared to other approaches in the literature.

جامعة الموصل

كلية علوم الحاسوب والرياضيات

كشف الوثائق المزيفة بأستخدام التحليل الطيفي: اعتماد نموذج شبكة

علاء امجد صالح

رسالة ماجستير

علوم الحاسوب

بأشراف

د. باسم محمد محمود

أستاذ مساعد

١٤٤٣ هـ

٢٠٢١ م

الخلاصة

يشهد عصرنا التكنولوجي الحالي ثورة كبيرة في مجال تطوير تطبيقات معالجة الوثائق. ان هذه التطبيقات يمكن استخدامها لاغراض مختلفة مثل تكوين وتحرير الوثائق. علاوة على ذلك، ان توفر الطابعات مثل الطابعات الليزرية والنقطية يجعل امكانية استخدامها متاحة للجميع وبدون قيود. يمكن استغلال هذه الطابعات والتطبيقات بشكل غير قانوني في التلاعب أو تقليد الوثائق الأصلية بواسطة المزورين.

هذه الرسالة تقترح إطارًا جديدًا غير خاضع للاشراف لكشف التزوير قادر على التمييز بين ما إذا كانت الوثيقة مزورة أو أصلية بناءً على التحليل الطيفي لحبر المستندات. يتم الحصول على اطياف احبار الوثائق (الأصلية والمشكوك فيها) بأستخدام تقنية التحليل الطيفي المستحثة بالليزر.

بعد ذلك، يتم حساب مصفوفة ارتباط لاطياف حبر الوثيقة الأصلية والمشكوك بها معًا. يتم بعدها تحويل مصفوفة الارتباط الى مصفوفة جوار لغرض تحويلها فيما بعد الى شبكة موزونة حسب مفاهيم نظرية المخطط. في الشبكة الموزونة المكونة، تمثل أطياف LIBS الرؤوس (V) للشبكة، وتمثل الارتباطات بين الأطياف أوزان (w) للحواف (E) بين الرؤوس المتقابلة. بعد تكوين الشبكة الموزونة للمستندات الأصلية والمشكوك بها، يتم تطبيق خوارزمية التجميع لوفيان. ومن ثم ، يتم قياس معدل الاكتشاف بناءً على عدد المجاميع (العناقيد) المتولدة في الشبكة.

تم اختبار النهج المقترح في إطار مجموعة متنوعة من الوثائق (مثل النصوص الرسمية والخطابات الرسمية) وأنواع مختلفة من الطابعات (على سبيل المثال ، نفث الحبر والليزر وآلات التصوير) بالإضافة إلى أنواع مختلفة من الأوراق. أظهرت النتائج ان الطريقة المقترحة توفر دقة بنسبة 95.6% في تمييز العينات المطبوعة من طابعات مختلفة باستخدام خوارزمية التجميع لوفاين. علاوة على ذلك، يوفر النهج المقترح معرفة نوع الحبر المستخدم في طباعة المستند المشكوك به وموديل الطابعة الذي تم استخدامه لطباعته. يتم أيضًا قياس النتائج التي تم الحصول عليها مع خوارزميات أخرى في الأدبيات. علاوة على ذلك ، يمكن تفسير نتائج النهج المقترح بصريًا ، وهو أمر مريح أكثر للمحققين الذين ليسوا محترفين في أنظمة الكمبيوتر. أخيرًا ، يعتبر النهج المقترح بسيطًا ولا يحتاج إلى حسابات معقدة مقارنة بالمناهج الأخرى في الأدبيات.