Library and Information Sciences
Print ISSN: 1680-9637
Online ISSN: 5977-2676

A Self-Recovery Digital Watermarking Approach for Tamper Detection of Handwritten and Printed Electronic Documents

Azadeh Bastani
*Corresponding author, Instructor, Department of Computer Engineering, Faculty of Engineering, Behbahan Khatam Alanbia University of Technology, Behbahan, Iran. E-mail: bastani@bkatu.ac.ir

Esmaeil Fatemi Behbahani
Assistant Professor, Department of Electrical Engineering, Faculty of Engineering, Behbahan Khatam alanbia University of Technology, Behbahan, Iran. E-mail: fatemi@bkatu.ac.ir

Abstract

Objective: The aim of this study is to present a brief discussion of the digital image watermarking for texts in order to prove the documents authentication and copyright protection and to introduce a new method of digital watermarking of printed and handwritten Persian documents image for tamper detection and recovery.

Methodology: This paper is based on an applied research. In this study, 20 Persian manuscripts and 35 printed texts were used to simulate the algorithm. In the proposed algorithm, using the information of the image pixels and combining it with a hidden key, we produced the digital watermark and embedded it in the text image. We then intentionally attacked the image and extracted the hidden watermark in the text. Finally, due to the self-recovery capability of the method, the tampered areas were detected and the document was restored to its original form.

Findings: In all 55 test images, separate watermark were embedded according to the proposed algorithm. To evaluate proposed system the three types of attacks includes adding, removing and updating, 30 random images were selected. In 20 images, all three types of manipulation were applied. Then the watermark was extracted and the results were studied. The accuracy of proposed method in detecting and correcting tampered areas was calculated according to two metrics: Tamper Detection Accuracy (TDA) and Tamper Recognition Accuracy (TRA). The algorithm was able to identify manipulated areas with an accuracy of 93.82 percent and correct and recover changes with an accuracy of 92.27 percent. It is important to note that the results obtained are based on the number of correctly identified bits. Therefore, the accuracy of less than 100 percent does not mean inability to detect some manipulated areas; Rather, as can be deduced from the intuitive results, this means that despite the correct diagnosis of the tampered

area, only some pixels of that area have not been correctly detected or recovered and therefore do not have a negative effect on the final tamper proofing. For imperceptibility test, PSNR value is estimated. The average value of PSNR in the suggested method is 40.18 dB.

Conclusion: Determining the documents authenticity without the availability of the original documents is one of the main necessities. Digital watermarking is one of the most powerful tools for document authentication and protection of copyright. Our proposed method has two special features that are considered to be its advantages over other methods. The first feature is the ability to work with all kinds of Persian text images. In addition to the ability to run on printed texts, it is also possible to model in handwritten texts as well as the combination of text and image. Another advantage is good rate of tamper detection and the ability to self-recovery of tampered images. The numerical and intuitive results indicate the efficiency of proposed scheme and so can be implemented for all handwritten and electronic printed texts. Therefore, this method can be used to protect the security of printed and handwritten documents in digital libraries, e-learning, e-government and any electronic document exchange system.

Keywords: Digital Watermarking, Document Images, Tamper Detection, Tamper Recovery.
Article type: Research

Central Library of Astan Quds Razavi
Library and Information Sciences, 2021, Vol. 24, No. 1, pp. 174-193.
Received: 20/07/2020 - Accepted: 10/10/2020

کتابداری و اطلاع رسانی
شاپا چاپی: ۹۶۳۷-۱۶۸۰
شاپا الکترونیکی: ۲۶۷۶-۵۹۷۷

طراحی یک روش الگوگذاری دیجیتال خود ترمیم جهت احراز اصالت اسناد چاپی و دستنویس الکترونیکی

آزاده باستانی
* نویسنده مسئول، مربی، گروه مهندسی کامپیوتر، دانشکده مهندسی، دانشگاه صنعتی خاتم الانبیاء بهبهان، بهبهان، ایران. رایانامه: bastani@bkatu.ac.ir

اسماعیل فاطمی بهبهانی
استادیارگروه مهندسی برق–الکترونیک، دانشکده مهندسی، دانشگاه صنعتی خاتم الانبیاء بهبهان، ایران. رایانامه: fatemi@bkatu.ac.ir

چکیده

هدف: پژوهش حاضر با هدف بررسی فعالیت‌های انجام شده در خصوص الگوگذاری دیجیتال اسناد و متون به منظور اثبات اصالت اسناد و حفاظت حق نشر و ارائه یک مدل جدید الگوگذاری دیجیتال متون فارسی چاپی و دستنویس با قابلیت تشخیص جعل سند و اصلاح تغییرات انجام شد.

روش شناسی: این پژوهش از نوع کاربردی است. برای شبیه‌سازی الگوریتم از ۲۰ متن دستنویس فارسی و ۳۵ متن چاپی استفاده شد. در الگوریتم پیشنهادی با استفاده از اطلاعات نقاط تصویر و ترکیب آن با یک کلید مخفی، الگوی دیجیتال تولید و در تصویر متنی ذخیره شد. سپس تصویر مورد حملات تعمدی قرار گرفته و الگوی پنهان شده در متن بازیابی گردید. در نهایت با توجه به قابلیت خود ترمیمی روش، نواحی جعل شده تصحیح و سند به شکل اولیه بازگردانده شد.

یافته‌ها: دقت روش پیشنهادی در تشخیص و تصحیح نواحی جعل شده بر حسب دو معیار دقت تشخیص (TDA) و دقت ترمیم (TRA) محاسبه شد. الگوریتم در حالت میانگین قادر به شناسایی نقاط دستکاری شده با دقت ۹۳/۸۲ درصد و تصحیح و بازیابی تغییرات با دقت ۹۲/۲۷ درصد بود. نامحسوس بودن الگو بر حسب مقیاس PSNR برابر با ۴۰/۱۸ دسی بل به دست آمد.

نتیجه‌گیری: نتایج حاصل از اندازه‌گیری عددی دقت مدل پیشنهادی و همچنین نتایج شهودی حاکی از کارایی مناسب آن است. الگوی ادغامی به صورت نامحسوس و با حفظ کیفیت بصری اسناد درج می‌گردد و برای تمامی متون دستنویس و چاپی الکترونیکی و برای کتابخانه‌های دیجیتال، آموزش الکترونیک و دولت الکترونیک قابل پیاده‌سازی و اجرا می‌باشد.

کلیدواژه‌ها: الگوگذاری دیجیتال، تصاویر متنی، تشخیص جعل سند، تصحیح جعل سند.

نوع مقاله: پژوهشی

کتابخانه مرکزی آستان قدس رضوی
کتابداری و اطلاع‌رسانی، ۱۴۰۰، دوره ۲۴، شماره ۱، شماره پیاپی ۹۳، صص. ۱۷۴–۱۹۳.
تاریخ ارسال: ۹۹/۳/۳۱ – تاریخ پذیرش: ۹۹/۷/۱۹

مقدمه

رشد روزافزون استفاده از شبکه‌های کامپیوتری و اینترنت، دسترسی سریع به منابع اطلاعاتی را آسان کرده است. مهمترین این منابع، انواع چند رسانه‌ای‌ها شامل متن، تصویر، صوت و فیلم هستند. دسترسی سریع و سهل به این رسانه‌ها، احتمال کپی‌برداری غیرمجاز و تغییرات تعمدی و غیرقانونی در آن‌ها را افزایش داده است. از طرفی تشخیص اصل یا جعل بودن آن‌ها بدون در دسترس بودن اصل سند کار راحتی نیست.

یکی از راه‌های حفظ امنیت، مخفی سازی اطلاعات[1] است که بر اساس هدف و نوع کاربرد، در دو دسته پنهان نگاری[2] و الگوگذاری دیجیتال[3] طبقه‌بندی می‌شوند. در پنهان نگاری، هدف اصلی مخفی سازی پیام و وجود هر نوع نشانه از پیام است و خود رسانه از اهمیت بالایی برخوردار نیست. اما در الگوگذاری، رسانه اصلی حائز اهمیت بوده و هدف مخفی سازی اطلاعاتی است که ضمن حفظ کیفیت رسانه، برای حفاظت حق نشر[4] و و یا تشخیص دستکاری آن قابل استفاده باشد. به همین دلیل در سال‌های اخیر الگوگذاری دیجیتال به صورت کاربردی مورد توجه واقع شده است. از جمله کاربردهای آن می‌توان به حفاظت حق چاپ، مدیریت و شناسایی محتوا، بایگانی اسناد، آموزش از راه دور، کتابخانه‌های دیجیتال، امضای دیجیتال، سیستم رأی‌گیری الکترونیکی، امنیت اطلاعات در دولت الکترونیک و... اشاره نمود.

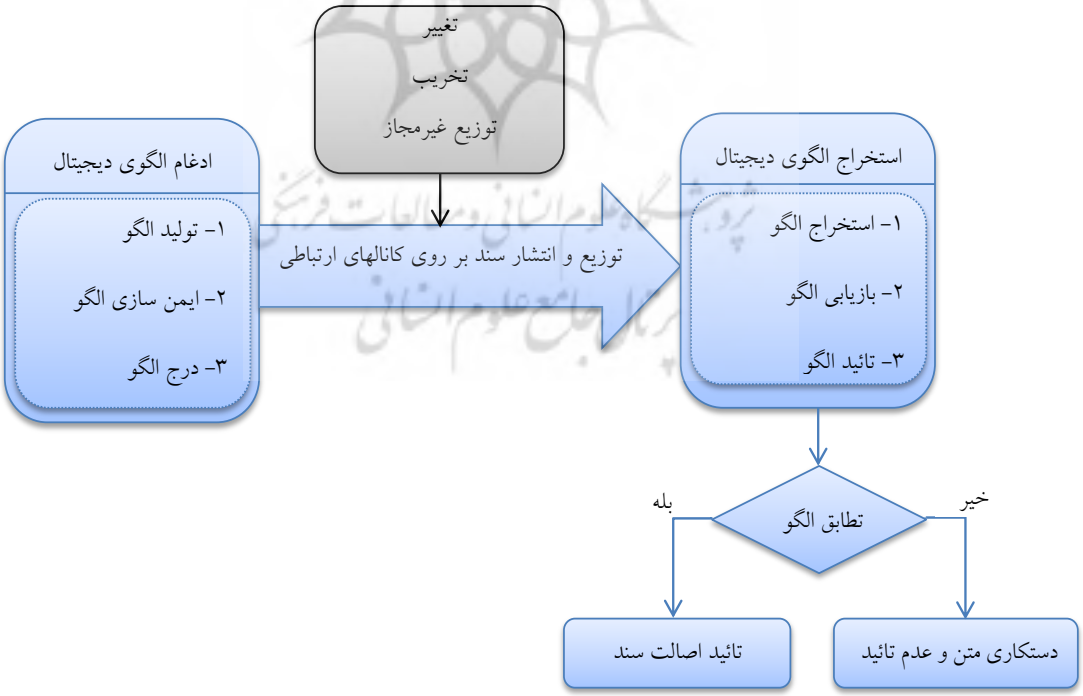
الگوگذاری دیجیتال یک فرآیند ادغام اطلاعات در رسانه اولیه و اصلی (میزبان) است به گونه‌ای که اطلاعات ادغامی تنها توسط یک شخص مجاز و معتبر قابل تشخیص باشد. بدین ترتیب که ابتدا یک الگوی اختصاصی شامل تصویر، شماره سریال، شماره شناسایی و... که نشان‌دهنده مالکیت رسانه میزبان است در هر سند درج شده، سپس جهت استفاده توزیع می‌گردد. در صورت بروز حملات دستکاری و جعل سند، در مرحله جداسازی الگو، دستکاری تشخیص داده شده و سند اصلاح می‌گردد. با چنین شیوه‌ای تمایل افراد به کپی‌برداری غیرقانونی کاهش می‌یابد. ادغام الگو می‌تواند به صورت قابل مشاهده و یا نامحسوس باشد. در نوع قابل مشاهده، نام مالک اثر، مهر و نشان تجاری یا لوگو مؤسسه در پس زمینه متن یا تصویر درج می‌گردد، در حالی که در انواع غیرقابل مشاهده ظاهر متن بدون تغییر باقی می‌ماند. در انواع نامحسوس، احتمال حذف یا تخریب الگو به دلیل غیرقابل مشاهده بودن آن کمتر است و به سبب امنیت بالاتر در سال‌های اخیر بیشتر مورد توجه قرار گرفته است. الگوگذاری دیجیتال نامحسوس بر اساس نوع کاربرد به دو دسته تقسیم‌بندی می‌شود:

1. Information Hiding
2. Steganography
3. Digital Watermarking
4. Copyright Protection

۱. الگوگذاری مقاوم: در این دسته الگوریتم‌ها، در صورت تغییر در تصویر و دستکاری عمدی، الگوی ادغامی در متن تغییر نمی‌کند. این دسته از الگوگذاری برای اثبات مالکیت اسناد و حفاظت حق نشر استفاده می‌شود. در این روش می‌توان به سند یک شماره شناسایی منحصر به فرد و غیرقابل دستکاری اعطاء نمود که در اثر کپی کردن غیرمجاز و یا تغییرات در محتوای متن از بین نرفته و قابل ردیابی باشد.

۲. الگوگذاری شکننده: در این نوع از الگوگذاری، الگوریتم‌های هوشمند مورد استفاده به گونه‌ای عمل ادغام الگو را انجام می‌دهند که در اثر تغییر در متن، الگو مخدوش یا تخریب شده و با توجه به میزان و مکان تخریب، می‌توان ناحیه دستکاری شده را تعیین نمود. برخی از این روش‌ها قابلیت خود ترمیمی نیز داشته و ناحیه تخریب شده را مجدداً بازیابی می‌نمایند. این نوع از الگوگذاری جهت اعتبارسنجی محتوای فایل‌های متنی تصویری و جلوگیری از جعل اسناد[1] قابل استفاده است.

همان گونه که در شکل ۱ نشان داده شده است یک مدل الگوگذاری دیجیتال، شامل دو بخش ادغام الگو و استخراج الگوست. اولین مرحله مربوط به تولید الگویی است که برگرفته از نام و مشخصات و یا هر اطلاعاتی از سوی صاحب فایل باشد. این اطلاعات به فرم الگوی دودویی و رشته‌ای از بیت‌های صفر و یک تبدیل شده و بر اساس الگوریتم‌های مشخص ایمن‌سازی می‌شود. نهایتاً الگو به صورت نامحسوس در مکان‌هایی از تصویر به گونه‌ای ذخیره می‌شود که در برابر تغییرات احتمالی تا حد امکان مصون بماند.



شکل ۱. معماری الگوگذاری دیجیتال

1. Documents Tampering

پس از آن اسناد از طریق اینترنت، پست الکترونیکی، شبکه‌های اجتماعی و... در دسترس سایرین قرار می‌گیرند. اسناد منتشره که شامل انواع مقالات، کتاب، گواهی‌نامه‌ها و اخبار است، ممکن است مورد حملات تغییر، تخریب و یا کپی برداری غیرمجاز قرار بگیرند. لذا در مرحله دوم بایستی الگوی موجود در سند را با الگوریتمی متناسب با روش درج الگو، استخراج نمود و بر اساس تطابق آن با الگوی اصلی، صحت و اصالت متن منتشره را مورد ارزیابی قرار داد. اصالت سند نقش مهمی در اعتبار آن دارد. با توجه به رشد روزافزون ابزارهای نرم‌افزاری که امکان تغییرات تعمدی در اسناد را آسان نموده است، به کارگیری روشی که اصل بودن اسناد را تأیید کند حائز اهمیت است.

در حال حاضر در زمینه تشخیص اصالت اسناد فارسی الکترونیکی و تصحیح آن‌ها، حفاظت حق چاپ و مالکیت معنوی متون روش‌های معدودی ارائه شده است و نیاز به انجام فعالیت در این زمینه و ارائه راهکارهای تحقیقاتی و اجرایی برای حل این مسئله احساس می‌شود. لذا هدف اصلی در این پژوهش ارائه یک راهکار علمی و عملی برای احراز اصالت اسناد چاپی و دستنویس الکترونیکی فارسی است. بدین منظور از یک روش الگوگذاری آماری مبتنی بر تصاویر متنی استفاده شده است که علاوه بر کاربرد در تأیید یا رد اصالت اسناد الکترونیکی، امکان تشخیص دستکاری‌های عمدی و تصحیح آن‌ها را فراهم می‌نماید. بدین صورت که از یک روش الگوگذاری شکننده استفاده می‌شود که برای تولید الگو، از اطلاعات مالک سند، اطلاعات کلی فایل سند و شدت رنگ[1] نقاط تصویر بهره می‌جوید. الگوگذاری به گونه‌ای انجام می‌شود که ضمن نامحسوس بودن، در اثر تغییرات یا حملات تعمدی به سند، الگوی مذکور تغییر کرده و با توجه به ویژگی‌های حاصل از تخریب، ناحیه دستکاری شده قابل شناسایی و اصلاح است.

روش ارائه شده را از جهات زیر مورد بررسی قرار می‌دهیم:

۱. آیا می‌توان بدون تغییر قابل ملاحظه و محسوس در ظاهر و کیفیت اسناد، الگویی در آن‌ها درج نمود؟
۲. آیا این روش توانایی تشخیص جعل اسناد را دارد؟
۳. کارایی این روش در ترمیم نواحی دستکاری شده چقدر است؟

پیشینه پژوهش

پنهان نگاری به عنوان یک هنر، از قدیمی ترین فنونی است که مورد توجه بوده است. در حدود سال‌های ۴۴۰ قبل از میلاد، حاکم یونانی که به دست داریوش زندانی شده بود، برای ارسال پیام‌های مخفی خود به

1. Pixel Intensity

لشکریانش، سر برده‌هایش را می‌تراشید، پیام‌ها را روی سر آن‌ها خالکوبی می‌کرد و بعد از رشد مجدد موها، برده‌ها را برای ارسال پیام اعزام می‌نمود. استفاده از جوهر نامرئی، کد کردن پیام‌ها در موزیک و... از سایر روش‌های مورد استفاده در گذشته بوده است.

اولین مطالعات علمی الگوگذاری متون با هدف حفاظت حق نشر و مخفی سازی اطلاعات به سال ۱۹۹۷ برمی‌گردد. از آن زمان، پژوهش در این زمینه به طور جدی ادامه داشته است. تحقیقات انجام شده را می‌توان در سه رویکرد زبانی، ساختاری و مبتنی بر تصویر دسته‌بندی نمود.

۱. رویکردهای زبانی: در این نوع از فعالیت‌ها که شاخه‌ای از علم پردازش زبان طبیعی است، بر اساس معنی کلمات و ساختار نحوی زبان کار می‌شود. الگوی واترمارک با استفاده از جایگزینی کلمات هم معنی، جابجایی قیدهای جمله، تبدیل کردن جملات معلوم و مجهول به یکدیگر و با حفظ ساختار درست جملات درج می‌گردد. یکی از کارهای اولیه در این زمینه توسط (آتالا، ۲۰۰۳) انجام شد که از درخت نحو برای یافتن کلمات هم معنی و درج واترمارک استفاده کرده است. از آن زمان تاکنون روش‌های بیشتری بر اساس نحو و قواعد زبان معرفی شده است، اما به دلیل دشواری الگوریتم‌های پردازش زبان طبیعی و عدم قطعیت روش‌های معرفی شده و همچنین عدم امکان استفاده از این روش‌ها در تمامی انواع اسناد، بازدهی کمی دارند.

۲. رویکردهای ساختاری: در این دسته از پژوهش‌ها، الگو به صورت یک رشته عددی مبنای دو، کد شده و در متن درج می‌گردد. در این روش‌ها مکان کلمات، حروف و نقاط و شکل ظاهری و نحوه نوشتن آن‌ها برای مخفی سازی تغییر می‌کند. بدین ترتیب که برای درج رشته‌ای از صفر و یک که همان بردار الگو است، بیت یک را با تغییرات بسیارکم و غیرقابل تشخیص در شکل فونت یا اندازه آن، جایگزین می‌نمایند و برای درج بیت صفر تغییری در شکل حروف و متن ایجاد نمی‌شود. داورزنی و یغمایی (۲۰۰۹) با استفاده از تغییر جزئی در میزان شیب چهار حرف «ر، ز، ژ، و» بیت‌های صفر و یک الگو را در متن جاگذاری می‌نمایند. شیرعلی (۲۰۰۶) با جابه‌جایی اندک نقطه‌های حروف به سمت بالا، بیت‌های الگو را در متن درج می‌نماید. یزدانی، دوستاری و یزدانی (۲۰۱۳) از حروف «ج، چ، ح، خ» استفاده می‌کنند. بدین ترتیب که برای درج بیت یک، مکان این حروف نسبت به خط زمینه کمی تغییر می‌کند و درج بیت صفر تغییری در آن‌ها نمی‌دهد. صفدری و لطیف (۱۳۹۵)، برای ادغام بیت‌های الگو از حروف دو نقطه‌ای و سه نقطه‌ای استفاده می‌کنند. بدین صورت که تغییر جزئی در فاصله بین نقاط، معرف بیت یک بوده و بیت صفر تغییری در مکان نسبی نقاط ایجاد نمی‌کند. در پژوهشی دیگر برای درج بیت یک از حروف و کلماتی که می‌توان به فرم کشیده نوشت مانند «رسول، محمد» استفاده شده است (القناهی، کبیر و تایتان، ۲۰۱۳). یزدانی و فلاح خورسند (۱۳۹۶) از لبه‌های منحنی دو حرف «ک، گ» برای ادغام الگو استفاده می‌کنند. در برخی از فعالیت‌ها از جابجایی مکان

خط، کلمات یا حروف منفرد استفاده شده است (کیم، مون ۲۰۰۳؛ یو، لیو ۲۰۰۹؛ العطار ۲۰۰۴؛ خدامی، یغمایی ۲۰۰۶). تمامی این روش‌ها بر روی متون چاپی قابل استفاده بوده ولی برای متن‌های دستنویس قابل اجرا نیستند. همچنین در متن‌هایی که همزمان دارای بخش‌های تصویری بوده و یا فرمت تصویری یا پی‌دی‌اف[1] داشته باشند کارایی ندارند. از سوی دیگر با تایپ مجدد متن، کپی متن، تغییر فونت و حذف یا تغییر در بخش کوچکی از فایل، الگوی درج شده تخریب می‌شود به گونه‌ای که الگوریتم‌های مورد استفاده قدرت بازیابی الگو و متن تغییر یافته را ندارند.

۳. رویکردهای مبتنی بر اطلاعات تصویر متن: در این دسته، متن به صورت تصویری و در فرمت‌هایی مانند JPG، BMP و PDF مورد استفاده قرار می‌گیرد. الگو می‌تواند به صورت یک متن، شماره سریال یا تصویر لوگو در متن گنجانده شود. در نوع شکننده، که برای تشخیص دستکاری به کار می‌رود، الگو ابتدا به فرم یک رشته متنی مبنای دو تبدیل و در تصویر اصلی مطابق با الگوی مشخص پنهان می‌گردد. روش کار بدین صورت است که اطلاعاتی از هر ناحیه از تصویر متنی به همراه بخشی از الگو در قسمت دیگری از تصویر جایگذاری می‌شود. با دستکاری هر ناحیه، به دلیل عدم تطبیق اطلاعات با ناحیه پشتیبان، تغییر متن تشخیص داده شده و اطلاعات اولیه از ناحیه پشتیبان اخذ و بازیابی می‌شود. به چنین الگوریتم‌هایی که توانایی اصلاح بخش‌های تخریب شده را دارند، خود ترمیم[2] گویند. کار در این حیطه به دلیل کارایی خوب و تنوع کاربرد، در دهه اخیر مورد توجه بسیار زیاد بوده است (باستانی ۱۳۹۸؛ کومار ۲۰۲۰؛ چتان، نیرمالا ۲۰۱۸؛ الحاج، فرفورا ۲۰۱۹؛ لوامر، تایان ۲۰۱۸). اکثر این فعالیت‌ها تنها بر روی تصاویر عادی اجرا شده و به دلیل حساسیت و لزوم دقت بالا در تصاویر متنی، تنها تعداد اندکی به طور خاص بر روی تصاویر متنی چاپی یا دستنویس تمرکز داشته‌اند. همچنین تحقیقات انجام شده در این دسته از الگوریتم‌ها بیشتر برای متن‌های انگلیسی و چینی به کار گرفته شده است. تفاوت همه این روش‌ها در نحوه استخراج اطلاعات نواحی و چگونگی نگاشت الگو در تصویر است. در (چتان، نیرمالا ۲۰۱۷) از ضرائب تبدیل کانتورلت پس از دو مرحله برای استخراج اطلاعات تصویر استفاده شده است. چتان، شیوانادا (۲۰۱۴) از تکنیک LSB[3] برای جایگذاری داده‌های پنهان استفاده کرده‌اند. الاحمد، الشیخلی و الدویخ (۲۰۱۳) با استفاده از تبدیل کسینوسی اقدام به درج الگو در فایل‌های PDF قرآن مجید نموده‌اند. این روش توانایی کار با تصاویر رنگی و خاکستری را دارد. یکی از کارها برای متن فارسی توسط (دارایی و مظفری، ۲۰۱۳) معرفی شده است که با استفاده از کدهای فراکتال الگوی واترمارک را در تصویر پنهان می‌کنند. این روش گرچه از کارایی خوبی برخوردار است اما به

1. PDF
2. Self-Recovery
3. Least Significant Bit

دلیل حجم بالای محاسبات فراکتال، بر روی تصاویر سیاه و سفید اجرا شده است. مشوش و دانیالی (۱۳۹۱) با استفاده از الگوریتم بایاس ضرایب و تبدیل موجک پنهان نگاری را انجام داده‌اند.

مروری بر پژوهش‌های انجام گرفته نشان می‌دهد که از بین سه رویکرد موجود، رویکرد سوم به دلیل تنوع بیشتر در تکنیک‌های علمی قابل استفاده، مقاومت و پایداری بالاتر در برابر انواع حملات و غیرقابل مشاهده بودن الگوی ادغام شده نسبت به دو رویکرد دیگر ارجحیت دارد. همچنین رویکرد سوم در مورد متون فارسی کمتر استفاده شده و به دلیل نوپا بودن، زمینه تحقیقاتی مناسبی می‌باشد. در پژوهش حاضر نیز روشی بر پایه رویکرد سوم ارائه گردیده است.

روش‌شناسی

بررسی مقالات مجلات علمی و کنفرانس‌های معتبر بین‌المللی ارائه شده در زمینه تشخیص جعل اسناد نشان داد که تعداد تصاویر مورد آزمایش در آن‌ها از ۳ تا ۶۰ تصویر متفاوت بوده است. در این پژوهش نیز به منظور ارزیابی جامع‌تر روش پیشنهادی، ۵۵ سند تصویری مدنظر قرار گرفت.

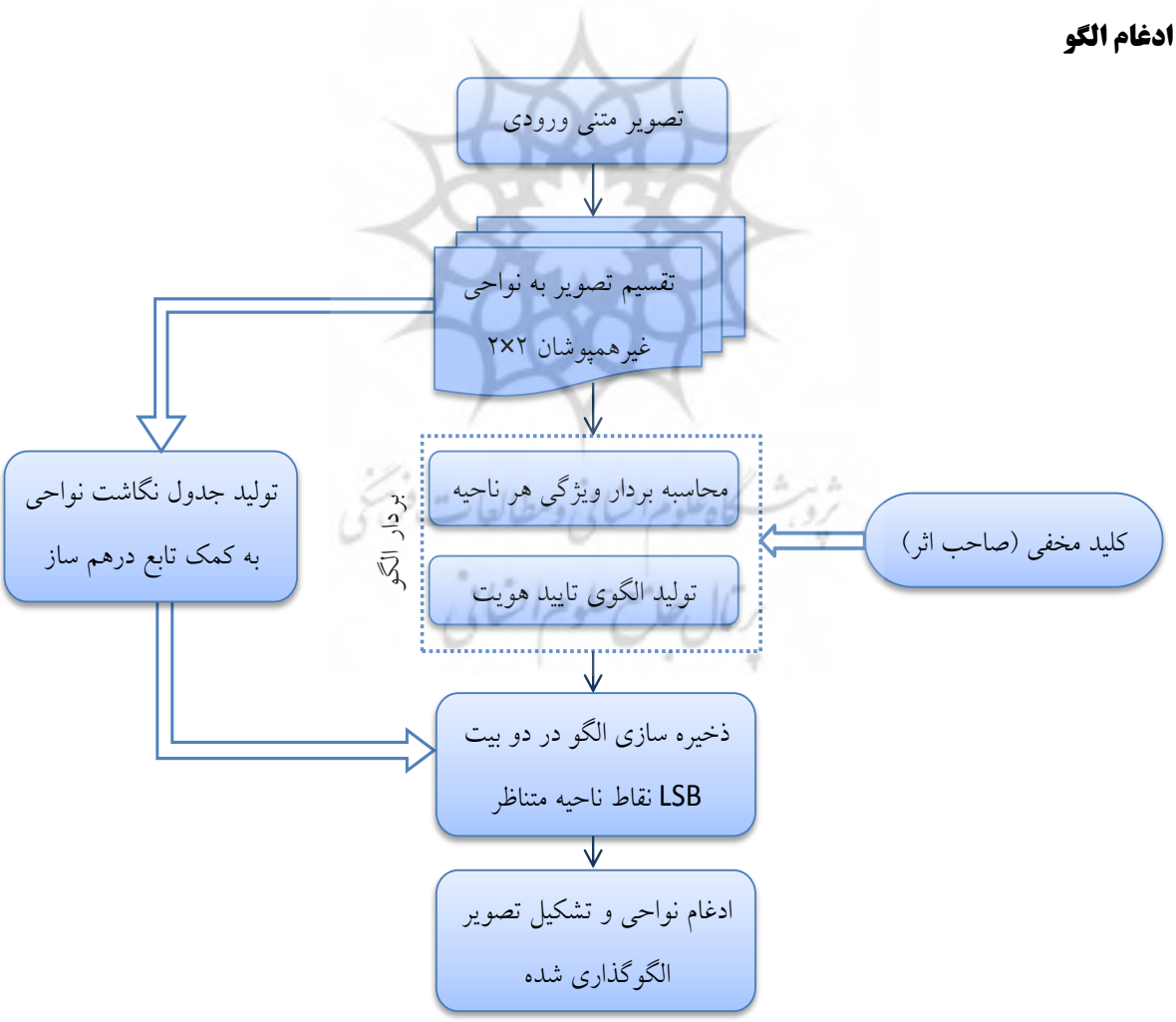
از طرفی بانک تصاویر اسناد فارسی معدودی وجود دارد که صرفاً جهت تشخیص متن و کلمه و کاربردهای OCR[1] کارایی داشته و معیارهای لازم برای اعتبارسنجی الگوریتم‌های تشخیص جعل اسناد را ندارند. در تهیه بانک اسناد مناسب برای شناسایی جعل بایستی پیچیدگی‌های لازم و تمامی جوانب احتمال جعل را در نظر گرفت. بنابراین، برای ارزیابی روش پیشنهادی با توجه به عدم وجود بانک تصاویر اسناد و مدارک فارسی مناسب برای بررسی جعل سند، اقدام به تهیه و جمع‌آوری اسناد شد. این اسناد به گونه‌ای انتخاب شدند که ویژگی‌های لازم جهت بررسی کارایی الگوریتم را دارا باشند. این ویژگی‌ها شامل تنوع در ضخامت، سایز و قلم[2] متن، فاصله بین کلمات، خطوط و پاراگراف، وجود همزمان متن چاپی و دستنویس، تنوع در محتوا و احتمال وجود بخش‌های تصویری غیرنوشتاری مانند تصویر افراد، امضاء، مهر و... در یک سند هستند. با توجه به موارد ذکر شده، در تهیه اسناد دستنویس از افراد مختلف با دستخط‌های متفاوت بهره جسته شد. همچنین در اسناد چاپی نیز، در هر سند، قلم‌های متفاوت با سایز و فاصله خطوط مختلف در نظر گرفته شد. تعدادی از اسناد نیز از طریق جستجو در موتور کاوش گوگل از بین اسناد و مدارک اداری جمع‌آوری و پس از بررسی اولیه انتخاب شدند. مدارک انتخابی به طور همزمان دارای متن چاپی، دستنویس، عکس، امضاء و... بوده و از نظر کیفیت و ابعاد تصویر، شرایط مناسب را دارا بودند.

1.Optical Character Recognition
2. Font

در مجموع ۲۰ سند دستنویس و ۲۰ متن چاپی، تهیه و اسکن شد. این اسناد شامل فاکتور فروش، کارت‌های شناسایی، گواهینامه‌های علمی، آموزشی و متون دستنویس با قلم و سایز متفاوت بودند. همچنین ۱۵ سند نیز از بین نامه‌های اداری و گواهینامه‌های منتشر شده در وبگاه‌های مراکز دولتی و خصوصی جمع‌آوری شدند. در نهایت برای یکسان‌سازی، تمامی تصاویر به ابعاد ۵۱۲×۵۱۲ نقطه تغییر سایز داده شد. شبیه سازی الگوریتم در محیط نرم‌افزار مهندسی متلب 2017 و با استفاده از ابزارهای پردازش تصویر و با پردازنده Intel Core i7 (2.8GHz) انجام گردید.

روش الگوگذاری شکننده خود ترمیم پیشنهادی، شامل دو مرحله ادغام الگو و استخراج الگو بوده که در ادامه تشریح شده است.

ادغام الگو



شکل ۲. مراحل درج الگو در تصویر

شکل ۲ مراحل ادغام الگو را نشان می‌دهد. ابتدا تصویر متن ورودی به چند ناحیه[۱] کوچکتر غیرهمپوشان[۲] و با ابعاد ۲×۲ نقطه[۳] تقسیم شد. سپس بردار الگوی هشت بیتی متشکل از دو بخش داده مخفی دو بیتی و اطلاعات ترمیم ناحیه شش بیتی برای هر یک از نواحی ایجاد گردید. برای تولید مقادیر بخش ترمیم، با میانگین‌گیری از چهار نقطه هر بلوک، یک بردار ویژگی محاسبه شد. برای ساخت بخش داده مخفی، یک شماره شناسایی بیانگر مالکیت صاحب اثر تعیین گردید. این اطلاعات مالک که کلید مخفی نامیده می‌شود به فرم دودویی[۴] تبدیل شد. برای بالا بردن امنیت سیستم، به کمک یک روال ابتکاری با توجه به مقدار کلید مخفی، برای هر ناحیه دو بیت تأیید هویت استخراج و با ترکیب آن با بردار ویژگی ناحیه، بردار الگو[۵] تولید شد. الگوی واترمارک تولید شده می‌بایست در ناحیه مربوطه ذخیره می‌گردید؛ اما در چنین شرایطی در صورت دستکاری آن ناحیه، الگو تخریب و بازسازی ناحیه متن نیز امکان‌پذیر نبود. بنابراین برای حل این مشکل و بالا بردن امنیت روش، اطلاعات هر ناحیه را در ناحیه دیگری از متن جایگزین کرده و برای انتخاب ناحیه کاندید هر بلوک، از یک تابع درهم ساز[۶] و یک دنباله عددی شبه تصادفی[۷] بهره گرفته شد. این بردار الگوی هشت بیتی در دو بیت کم ارزش[۸] شدت رنگ چهار نقطه از ناحیه متناظر با هر ناحیه جایگزین شد. با تکرار این کار برای تمام نواحی و ترکیب مجدد نواحی، تصویر الگوگذاری شده به دست آمد.

استخراج الگو

مراحل استخراج الگو در شکل ۳ نشان داده شده است. برای استخراج الگو از تصویر متنی، تصویر مجدداً به همان نواحی ۲×۲ مشابه مرحله قبل تقسیم گردید و با استفاده از دنباله شبه تصادفی و تابع درهم ساز مرحله ادغام، نگاشت هر ناحیه معلوم گردیده و بردار الگوی هر بخش مجدداً محاسبه شد. این بردار هشت بیتی با مقادیر ذخیره شده در ناحیه نگاشت شده متناظر هر بلوک و طبق محاسبات ریاضی تطبیق داده شد و نواحی ناسازگار تعیین گردید. به عنوان مثال، در صورت دستکاری ناحیه A و نظر به این که اطلاعات اصلی آن در مرحله ادغام در ناحیه B ذخیره شده بود، در بلوک B عدم تطبیق رخ داده و ناحیه تغییریافته مشخص گردید. در گام بعدی، تمامی نواحی پس از محاسبه میزان تطبیق با ناحیه نگاشت شده متناظر و اندازه‌گیری احتمال تخریب به دو دسته تأیید شده و دستکاری شده علامت‌گذاری شدند. نواحی تأیید شده نیاز به تغییر و اصلاح

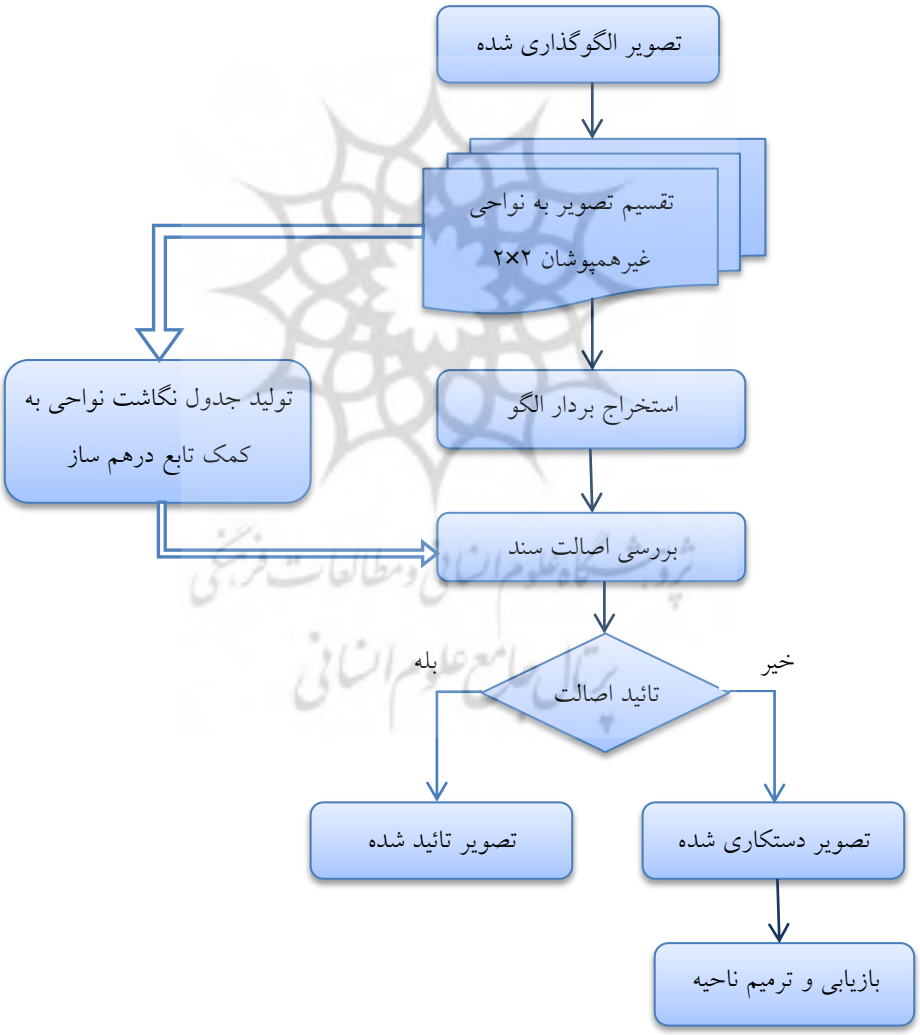
1. Block
2. Non-overlapping
3. pixel
4. Binary
5. Watermark
6. Hashing Function
7. Pseudo Random Sequence
8. LSB

نداشتند، در حالی که در نواحی دستکاری شده، می‌بایست اطلاعات قبلی بازیابی می‌شد. این اطلاعات در مرحله ادغام در ناحیه نگاشت شده متناظر و در قالب شش بیت اطلاعات ترمیم، گنجانده شده بود که پس از بازخوانی، جایگزین پیکسل‌های ناحیه تغییر یافته شد.

کارایی روش الگوگذاری شکننده پیشنهادی با استفاده از دو معیار دقت تشخیص جعل[1] (TDA) و دقت ترمیم[2] (TRA) سنجیده شده است. این مقادیر به صورت زیر محاسبه می‌گردند:

$$TDA = \frac{\text{تعداد بیت‌های دستکاری شده درست تشخیص داده شده}}{\text{تعداد کل بیت‌های دستکاری شده}}$$

$$TRA = \frac{\text{تعداد بیت‌های درست ترمیم شده}}{\text{تعداد کل بیت‌های دستکاری شده}}$$



شکل۳. مراحل استخراج الگو و بررسی اصالت سند

1. Tamper Detection Accuracy
2. Tamper Recovery Accuracy

برای ارزیابی روش پیشنهادی در مجموعه تصاویر متنی مورد آزمایش، الگوهای دیجیتال تصادفی درج شد. سپس سه نوع از متداول‌ترین روش‌های جعل سند که شامل الحاق، امحاء و قلم بردن در متن است، بر روی آن‌ها اعمال شد. پس از انجام مراحل استخراج الگو، نواحی دستکاری شده تشخیص و اطلاعات پیش از جعل، بازیابی و نواحی ترمیم شدند. میزان دقت و کارایی روش با استفاده از دو معیار TDA و TRA مورد بررسی قرار گرفت.

یافته‌ها

در ابتدا برای هر یک از ۵۵ تصویر اولیه، الگوهای دیجیتال تصادفی مجزا تولید و مطابق با روش تشریح شده در بخش ادغام الگو به عنوان کلید مخفی در تصاویر گنجانده شد. شایان ذکر است که درج اطلاعات مالک اثر نیز به جای الگوهای تصادفی، امکان‌پذیر است.

نکته حائز اهمیت این است که پس از درج الگو، شدت رنگ نقاط تصویر دچار تغییر رنگ جزئی می‌گردد. برای بررسی میزان غیرقابل مشاهده و نامحسوس بودن الگوی ادغامی از معیاری به نام PSNR[1] استفاده می‌شود. حداکثر مقدار PSNR که بر حسب واحد دسی بل محاسبه می‌شود، بسته به بالاترین سطح رنگ تصویر، ابعاد تصویر و سایز الگوی ادغامی متغیر است. بزرگ بودن PSNR نشان از میزان شباهت تصویر اولیه و تصویر الگوگذاری شده و در نتیجه غیر قابل کشف بودن الگو دارد. PSNR به صورت زیر تعریف می‌شود:

$$\mathrm{PSNR} = 10 \log \frac{255^2}{\mathrm{MSE}}$$ رابطه ۱)

که در آن MSE[2] میانگین مربعات خطا بوده و به روش زیر محاسبه می‌شود:

$$MSE = \frac{1}{X*Y} \sum_{i=1}^{X} \sum_{j=1}^{Y} (I_{ij} - W_{ij})^2$$ رابطه ۲)

در این رابطه I_{ij} پیکسل‌های تصویر اصلی به سایز X*Y بوده و W_{ij} پیکسل‌های تصویر الگوگذاری شده با همان سایز X*Y می‌باشد. به طور معمول تصاویر الگوگذاری شده با مقدار PSNR بزرگتر از ۲۸ دسی بل قابل قبول هستند (سینگ، داو و موهان ۲۰۱۴). در روش پیشنهادی پس از درج الگو در اسناد مورد آزمایش، مقدار میانگین PSNR برابر با ۴۰/۱۸ دسی بل به دست آمد که بیانگر نامحسوس بودن الگو و تطابق مطلوب تصاویر اولیه با تصاویر جدید است. با توجه به این عدد، می‌توان چنین استنباط نمود که الگوریتم به کار گرفته شده با وجود درج الگو در اسناد، کیفیت بصری آن‌ها را تغییر نمی‌دهد و سند از دید بیننده، تغییر قابل مشاهده‌ای

1. Peak Signal to Noise Ratio
2. Mean Square Error

نیافته است و جاعلان احتمالی به وجود الگوی مخفی مشکوک نمی‌شوند. بنابراین در پاسخ به پرسش اول پژوهش می‌توان گفت با توجه به نرخ PSNR روش حاضر قابلیت درج الگو با حفظ کیفیت ظاهری فایل تصویری اسناد را دارد.

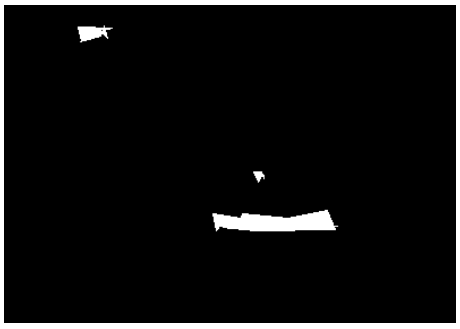
شکل ۴ یک نمونه از تصویر متنی مورد استفاده، که الگوی واترمارک نامحسوس در آن درج شده است را نشان می‌دهد. این سند یک گواهی حضور در دوره آموزشی است. دلیل انتخاب این نمونه از اسناد، پیچیدگی‌های لازم نظیر وجود همزمان متن چاپی و دستنویس، استفاده از متن با قلم‌های متفاوت، ضخامت و سایزهای متنوع، فواصل مختلف خطوط و وجود شکل و امضاء و... است.

در شکل ۵ سه نوع حمله افزودن، حذف و تغییر بر روی تصویر شکل ۴ اعمال شده است. به این ترتیب که در سند، شماره نامه حذف شده، میزان ساعات برگزاری دوره تغییر داده شده و عبارت «موفق به کسب نمره ۱۰۰ از ۱۰۰ شده‌اند» به سند اضافه شده است. این موارد در شکل به طور مشخص نشان داده شده است.


گواهی حضور در دوره آموزشی

شکل ۴. نمونه‌ای از تصویر الگوگذاری شده

شکل ۵. تصویر دست کاری شده شکل ۴



شکل ۶. نواحی نامعتبر تشخیص داده شده توسط الگوریتم

با اعمال الگوریتم پیشنهادی بر روی تصویر شکل ۵، نواحی معتبر از نقاط تأیید نشده تفکیک می‌شوند. شکل ۶ مناطقی که توسط الگوریتم به عنوان نواحی نامعتبر شناخته شده است را با رنگ سفید نشان می‌دهد. با توجه به این شکل مشاهده می‌شود که نواحی غیرمجاز یافته شده توسط الگوریتم پیشنهادی، به درستی تشخیص داده شده و روش مورد استفاده از دقت و امنیت لازم برخوردار است.

برای ارزیابی روش معرفی شده از لحاظ نرخ تشخیص دستکاری و ترمیم اسناد، تمامی ۵۵ تصویر الگوگذاری شده مورد آزمایش، در معرض سه نوع حمله الحاق (افزودن)، امحاء (حذف) و قلم بردن (تغییر متن) قرار داده شدند. برای هر یک از سه مدل حمله، ۳۰ تصویر به طور تصادفی از مجموعه تصاویر آزمایشی انتخاب و استفاده شد. بدین ترتیب که برای آزمایش حمله حذف، بخش‌هایی از هر یک از تصاویر حذف و در حمله قلم بردن، قسمت‌هایی از سند تغییر داده شد. در جعل از نوع افزودن نیز به قسمت‌هایی از تصاویر متون جدید اضافه شد. همچنین در ۲۰ تصویر هر سه نوع دستکاری به طور همزمان اعمال شد.

جدول ۱. نتایج الگوریتم پیشنهادی بر حسب میزان تشخیص و تصحیح متن

نوع جعل (دستکاری)	TDA (درصد تشخیص جعل)	TRA (درصد تصحیح جعل)
الحاق (اضافه کردن)	۹۴/۷۶	۹۳/۳۱
امحاء (حذف کردن)	۹۳/۴۹	۹۱/۸۸
قلم بردن (تغییر دادن)	۹۴/۲۱	۹۲/۶۸
ترکیبی از هر سه نوع	۹۲/۸۳	۹۱/۲۵

جدول ۱ نتایج اعمال الگوریتم پیشنهادی بر روی بانک تصاویر متنی دستنویس و چاپی بر حسب میانگین مقادیر TDA و TRA را برای انواع حملات دستکاری متن نشان می‌دهد. بر اساس نتایج به دست آمده در روش پیشنهادی، نرخ تشخیص نواحی دستکاری شده در حملات از نوع الحاق با مقدار ۹۴/۷۶ درصد

نسبت به دو نوع دیگر دستکاری از دقت بالاتری برخوردار است. همچنین در مقایسه با سایر حملات، کمترین نرخ تشخیص با میزان ۹۳/۴۹ درصد مربوط به حمله امحاء می‌باشد. در حالتی که هر سه نوع حمله به صورت ترکیبی به اسناد اعمال شده است، نرخ تشخیص با وجود کمتر بودن نسبت به حملات تکی، همچنان از دقت مطلوب ۹۲/۸۳ درصد برخوردار است.

در خصوص نرخ تصحیح نواحی دستکاری شده، با توجه به پایین‌تر بودن درصد تصحیح نسبت به نرخ تشخیص، این نکته قابل استنباط است که بخش اعظم نواحی تغییر داده شده به طرز درستی تصحیح شده‌اند و تنها درصد کمی از نقاط، با وجود تشخیص توسط الگوریتم، به نحو دقیق قابل بازیابی نبوده است. بیشترین نرخ میانگین تصحیح، برای حمله الحاق و برابر با ۹۳/۳۱ درصد و کمترین میزان برای دستکاری ترکیبی با دقت ۹۱/۲۵ درصد به دست آمده است. نکته قابل توجه این است که این دو معیار، درصد تشخیص درست نواحی جعل شده و تصحیح آن نواحی را بر حسب بیت محاسبه می‌نمایند؛ بنابراین دقت کمتر از ۱۰۰ درصد به معنای عدم توانایی در تشخیص برخی از نواحی نیست؛ بلکه آنچنان که از نتایج شهودی نیز قابل استنباط است، بدین معناست که با وجود تشخیص درست ناحیه دستکاری شده، تنها برخی از نقاط آن ناحیه به درستی تشخیص یا بازیابی نشده‌اند و لذا در روند تشخیص نهایی جعل، نتیجه منفی تأثیرگذار ندارد. بنابراین در پاسخ به پرسش دوم و سوم می‌توان گفت که با توجه به مقادیر عددی به دست آمده برای میزان تشخیص و تصحیح بخش‌های دستکاری شده و همچنین نتایج بصری تشخیص که نمونه‌ای از آن در شکل‌های ۵ و ۶ آورده شد، پژوهش حاضر توانایی تشخیص مطلوب و تصحیح جعل در اسناد را دارد.

نتیجه‌گیری و پیشنهادها

در عصر دیجیتال ایجاد تصاویر و متون ساختگی با استفاده از ابزارهای نرم‌افزاری به سادگی امکان‌پذیر شده است. لذا تعیین اعتبار و صحت اسناد بدون در دسترس بودن اصل سند از نیازهای اصلی به شمار می‌آید. الگوگذاری دیجیتال یکی از ابزارهای قدرتمند برای صحت‌سنجی اسناد و حمایت از حق مالکیت اثر است. بیشترین فعالیت‌های آغازین انجام شده در زمینه الگوگذاری دیجیتال، برای اسناد چاپی بوده و بر شکل کلمات و حروف تمرکز داشتند. اشکال اصلی آن‌ها امنیت کم در برابر حملات جعل بود به گونه‌ای که با تغییر و دستکاری حروف متن و یا اضافه و حذف کردن بخشی از متن، امکان تشخیص و اصلاح در آن‌ها وجود نداشت. ضمن این که در این روش‌ها حوزه فعالیت تنها به اسناد کاملاً متنی و چاپی محدود بوده و برای اسنادی که مهر و امضاء، عکس و... در آن‌ها وجود داشته باشد استفاده کاربردی ندارد. در ادامه نوع جدیدی از روش‌ها مورد توجه قرار گرفت که برای رفع مشکلات قبلی، بر اطلاعات کل متن و شدت رنگ نقاط تصویر

استوار بود. اکثر این نوع پژوهش‌ها بر روی متون انگلیسی و چینی انجام شده است. پژوهش حاضر از اولین تحقیقات انجام شده با روش‌های آماری بر روی اسناد فارسی الکترونیکی می‌باشد. این روش دارای دو ویژگی خاص می‌باشد که از مزیت‌های آن نسبت به سایر روش‌ها محسوب می‌گردد. ویژگی اول قابلیت کار با انواع متون تصویری فارسی است. به گونه‌ای که علاوه بر قابلیت اجرا بر روی متون چاپی، در متون دستنویس و همچنین ترکیب متن و تصویر نیز امکان الگوگذاری وجود دارد. مزیت دیگر آن قابلیت خود اصلاحی تصاویر تغییریافته می‌باشد، بدین صورت که با درج اطلاعات آماری از نواحی مختلف سند، در صورت تغییر یافتن بخش‌هایی از آن، با بازیابی داده‌های ذخیره شده، آن نواحی مجدداً ترمیم می‌شوند. در این روش از ۲۰ سند دستنویس، ۲۰ سند چاپی و ۱۵ سند ترکیبی استفاده شد. هر یک از تصاویر به نواحی کوچکتر تقسیم شده و بر اساس یک الگوی بیتی تصادفی و میانگین شدت رنگ هر ناحیه بردار الگو ساخته و در سایر نواحی جایگزین شد. این تصاویر مورد دستکاری عمدی قرار گرفتند. سه نوع تغییری که در تصاویر داده شد شامل اضافه کردن متن، حذف کردن قسمتی از متن و تغییر دادن کلمات بود. سپس با روال معکوس مرحله ادغام، نواحی مورد حمله واقع شده شناسایی و با استفاده از اطلاعات ذخیره شده در تصویر، تغییرات اعمال شده بازیابی و اصلاح شد.

پژوهش حاضر تلاش نمود تا به کمک الگوریتم‌های پردازش تصویر و ابزارهای آماری و ریاضی روشی در جهت مقابله با جعل اسناد معرفی کند که کمترین تأثیر مخرب بر تصویر و کمترین خطا در شناسایی جعل داشته باشد. بر این اساس بازدهی الگوریتم پیشنهادی از سه منظر مورد بررسی قرار گرفت:

- ≠ میزان نامحسوس و غیرقابل کشف بودن الگوی ادغامی در اسناد الکترونیکی
- ≠ نرخ تشخیص نواحی دستکاری شده
- ≠ قدرت تصحیح نواحی جعل شده

عامل اول یعنی غیرقابل تشخیص بودن وجود الگو از نظر محاسباتی با معیار PSNR بررسی‌گردید. در روش ارائه شده در این پژوهش، مقدار PSNR برابر با ۴۰/۱۸ دسی بل بوده که نشان از نامحسوس بودن الگو دارد. این امر از نظر بصری و با مشاهده عینی تصاویر الگوگذاری شده نیز قابل تأیید بوده و هم‌راستا با نتایج عددی است. عامل دوم در کارایی روش، میزان تشخیص درست نواحی جعل شده است که با معیار عددی TDA سنجیده شد. میانگین مقدار TDA به دست آمده در این پژوهش معادل با ۹۳/۸۲ درصد است که تعداد نقاط جعلی درست تشخیص داده شده نسبت به کل نقاط تغییر یافته یا به عبارتی قابلیت شناسایی جعل را بیان می‌نماید. سومین عامل، توانایی روش معرفی شده در بازیابی سند اصلی بوده که با معیار عددی TRA بررسی و نتایج شبیه سازی نشان داد که روش پیشنهادی در حالت متوسط، با دقت ۹۲/۲۷ درصد قابلیت

بازیابی سند اصلی را دارد. علاوه بر این نتایج شهودی حاصل از اعمال الگوریتم پیشنهادی بر روی تصاویر مورد حمله واقع شده نشان داد که نواحی دستکاری شده به خوبی تشخیص داده شده و اصلاح می‌شوند. نتایج به دست آمده گواه کارایی قابل قبول روش پیشنهادی است، لذا از این روش می‌توان جهت حفظ امنیت اسناد چاپی و دستنویس در کتابخانه‌های دیجیتال، آموزش الکترونیک، دولت الکترونیک و هر نظام تبادل الکترونیکی اسناد بهره جست.

با توجه به این که پژوهش حاضر بر روی اسنادی که دارای فایل تصویری هستند اجرا شده است، جهت پژوهش‌های آینده پیشنهاد می‌شود نتایج بر روی اسناد از نوع PDF نیز بررسی گردد. بخش زیادی از اسناد الکترونیکی به صورت PDF ذخیره می‌گردند و این در حالی است که بدنه اصلی این فایل‌ها دارای ساختاری مشابه با فایل‌های تصویری است.

همچنین می‌توان سایر روش‌های آماری از جمله انواع تبدیلات حوزه فرکانس مانند تبدیل فوریه، موجک و... نیز در مرحله ادغام الگو مورد استفاده قرار گرفته و کارایی آن‌ها از نظر سرعت اجرا و دقت تشخیص جعل و تصحیح اسناد با روش پیشنهادی مقایسه گردد.

سپاسگزاری

نگارندگان بر خود لازم می‌دانند از داوران محترم به خاطر ارائه نظرهای ساختاری و علمی سپاسگزاری نمایند.

منابع

باستانی، آزاده (۱۳۹۸). الگو گذاری ایمن و با ظرفیت بالا در تصاویر دیجیتال با استفاده از گشتاورهای چبیشف، مقاله ارائه شده در *سومین کنفرانس بین المللی محاسبات نرم*، رودسر، دانشگاه گیلان.

صفدری، طاهره؛ لطیف، علی محمد (۱۳۹۵). ارائه یک روش جدید جهت پنهان نگاری اطلاعات در فایل متنی فارسی، *پژوهش‌های حفاظتی ـ امنیتی دانشگاه جامع امام حسین(ع)*، ۵(۱۸)، ۴۳-۵۹.

مشوش، زینب؛ دانیالی، حبیب الله (۱۳۹۱). نهان نگاری برگشت پذیر تصاویر دیجیتال با قابلیت حفاظت از کپی رایت و تشخیص تغییر تصویر نهان نگاری شده، مقاله ارائه شده در *نهمین کنفرانس بین‌المللی انجمن رمز ایران*، تبریز، انجمن رمز ایران.

یزدانی، وحید؛ موسی، فلاح خورسند (۱۳۹۶). مخفی کردن اطلاعات در تصاویر متنی با استفاده از پیکسل‌های موجود روی لبه‌های حروف فارسی، *دو ماهنامه نخبگان علوم و مهندسی*، ۲(۲).

References

Alahmad, M. A., Alshaikhli, I., & Alduwaikh, A. E. (2013). A New Fragile Digital Watermarking Technique for a PDF Digital Holy Quran, *International Conference on Advanced Computer Science Applications and Technologies*, Kuching, 250-253.

Alattar, A. M., & Alattar, O. M. (2004). Watermarking Electronic Text Documents Containing Justified Paragraphs and Irregular Line Spacing, *Proceedings of SPIE*, 5306, 685-695.

Alginahi, Y. M., Kabir N., & Tayan, O. (2013). An enhanced Kashida-based watermarking approach for Arabic text-documents, *International Conference on Electronics, Computer and Computation (ICECCO)*, Ankara, 301-304.

Al-Haj, A., & Farfoura, M. (2019). Providing Security for E-Government Document Images Using Digital Watermarking in the Frequency Domain, *5th International Conference on Information Management (ICIM)*, Cambridge, United Kingdom, 77-81.

Atallah M. J., Raskin, V., & Hempelmann, F. & et al. (2003). Natural Language Watermarking and Tamperproofing, *Petitcolas F.A.P. (eds) Information Hiding, Lecture Notes in Computer Science*, 2578, Springer, Berlin, Heidelberg

Bastani, A. (2019). Robust and high capacity digital image watermarking using chebychev moments, *International conference on soft computing*, Guilan, Rudsar. *(in Persian)*

Chetan, K. R., & Nirmala, S. (2017). An intelligent fragile watermarking scheme based on contourlets for effective detection, localization and recovery of tampered regions in handwritten document images, *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Udupi, 405-410.

Chetan, K. R., & Nirmala, S. (2018). Intelligent Multiple Watermarking Schemes for the Authentication and Tamper Recovery of Information in Document Image, *Choudhary R., Mandal J., Bhattacharyya D. (eds) Advanced Computing and Communication Technologie. Advances in Intelligent Systems and Computing*, 562, Springer, Singapore

Chetan, K. R., & Shivananda, N. (2014). A new fragile watermarking approach for tamper detection and recovery of document images, *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, New Delhi, 1494-1498.

Daraee, F., & Mozaffari, S. (2013). Watermarking in Binary Document Images using Fractal Codes, *Pattern Recognition Letters*, 35, 120-129.

Davarzani, R., & Yaghmaie, K. (2009). Farsi Text Watermarking Based on Character Coding, *International Conference on Signal Processing Systems*, Singapore, 152-156.

Khodami, A., & Yaghmaie, K. (2006). Persian Text Watermarking, *Advances in Multimedia Information Processing, Lecture Notes in Computer Science*, 4261, Springer, Berlin, Heidelberg, 927-934.

Kim, Y., Moon, K., & Oh, I. (2003). A Text Watermarking Algorithm based on Word Classification and Inter word Space Statistics, *Proceedings of the Seventh International Conference on Document Analysis and Recognition*, 775-779.

Kumar, A. (2020). A Review on Implementation of Digital Image Watermarking Techniques Using LSB and DWT, *Advances in Intelligent Systems and Computing*, 933,595-602

Kumar, S., Basant, K., Mayank, D., & Anand, M. (2015). Multiple watermarking on medical images using selective discrete wavelet transforms coefficients, *Journal of Medical Imaging and Health Informatics*, 5, 1–8.

Laouamer, L., & Tayan, O. (2018). Performance Evaluation of a Document Image Watermarking Approach With Enhanced Tamper Localization and Recovery, *IEEE Access*, 6, 26144-26166.

Moshavesh, Z., & Daniali, H. (2012). Reversible digital image steganography for copyright protection and tamper detection, *International Conference on Iranian Security Community*, available at https//www.civilica.com/Paper-ISCC09-ISCC09_031.html *(in Persian)*

Safdari, T. & Latif, A. (2016). A new method for Information Hiding in Persian text file, *Journal of Protection and Security*, 5(18), 43-59. *(in Persian)*

Singh, A. K., Dave, M., & Mohan, A. (2014). Hybrid Technique for Robust and Imperceptible Image Watermarking in DWT–DCT–SVD Domain, *Natl. Acad. Sci. Lett.* 37**,** 351–358.

Shirali-Shahreza, M., & Shirali-Shahreza, M. (2006). A New Approach to Persian/Arabic Text Steganography, *Proceedings of 5th IEEE/ACIS International Conference on Computer and Information Science*, 4(11), 1692-1698.

Yazdani, V., Doostari, M. A., & Yazdani, H. (2013). A New Method to Persian Text Watermarking Using Curvaceous Letters, *Journal of Basic and Applied Scientific Research*, 3(4), 125-131

Yazdani, V., & Fallah khorsand, M. (2017). A New Method to Persian Text Watermarking Using Curvaceous Letters, *Journal of Science and Engineering Elites*, 2(2). *(in Persian)*

Yu, Z., & Liu, X. (2009). A New Digital Watermarking Scheme Based on Text, *International Conference on Multimedia Information Networking and Security*, Hubei, 138-140.