

The 1th Conference on
**Optoelectronics, Applied Optics and Microelectronics
(OAM)**
Department of Advanced Technologies,
University of Mohaghegh Ardabili
Namin, Ardabil, Iran
Aug. 17-19, 2019

کاربرد تصویربرداری فراطیفی برای شناسایی جوهر در اسناد دستنویس

پیمان سلطانی^{*}، مجید پناهی، احمد درودی

گروه فیزیک، دانشکده علوم، دانشگاه زنجان، زنجان

شرکت پرتو افزار صنعت

چکیده: برای اجرای صحیح قانون در اسناد و مدارک معتبر، تشخیص جوهر تبدیل به یک عامل مهم برای شناسایی اسناد جعلی شده است. در این مقاله برای تجزیه و تحلیل جوهر در اسناد دستنویس و سنجش نوع خودکار، از یک رویکرد با استفاده از تصویربرداری فراطیفی (Hyperspectral Imaging) و طبقه‌بندی ماشین بردار پشتیبان (SVM) استفاده می‌شود. روش استفاده شده نتایج را از طریق مطالعه دقیق بر خواص جوهر و خودکار بدون تماس با نمونه به دست می‌دهد. ما از یک روش طبقه‌بندی مبتنی بر ویژگی‌های فراطیفی جوهر که از خواص فیزیکی آن مشتق شده است استفاده می‌کنیم. علاوه بر این، یک پایگاه داده از منحنی‌های فراطیفی چند نوع جوهر هم‌رنگ ساخته شده است که برای به دست آوردن خصوصیات جوهرهای مختلف استفاده می‌شود. روش پیشنهادی برای شناسایی جوهر با استفاده از هفت خودکار مختلف و بیش از ۲۰۰ نمونه مورد آزمایش قرار گرفته است و نتایج ارائه شده است.

کلید واژه: جوهر، تصویر فراطیفی، طبقه‌بندی، ماشین بردار پشتیبان (SVM)

Hyperspectral imaging for Ink Identification in Handwritten Documents

Peyman Soltani^{*}, Majid Panahi, Ahmad Darudi

- 1- Department of Physics, Faculty of Science, University of Zanjan, Zanjan
- 2- Parto Afzar Sanat corporation
^{*} pn.soltani@gmail.com

Abstract: In order to correctly enforce the law in valid documents, intrusion detection has become an important factor in identifying fake documents. In this paper, an approach using Hyperspectral Imaging and Support-vector machine classification (SVM) is used to ink Analysis in handwritten documents and pen type test. The method used gives the results through a careful study of the properties of the ink and the pen without touching the sample. We use a classifier based on the inherent properties of the ink that derives from its physical properties. In addition, a database of hyperspectral curves is made up of several ink types that are used to obtain the properties of different inks. The proposed method for Ink Identification using seven different pens and more than 200 samples has been tested and the results are presented.

Keyword: Ink, Hyperspectral Image, Classification, Support-Vector Machine (SVM)

۱- مقدمه

تفکیک پذیری بالای طیفی فراهم می‌کند. از آنجایی که هر عنصر طبیعی دارای خواص جذب و انعکاس بسته به ساختار اتمی آن است، هنگامی که یک جوهر با نور سفید روشن می‌شود، برخی از طول موج‌ها را جذب می‌کند و برخی را انعکاس می‌دهد. نور سفید حاوی طول موج‌های مختلف است و پاسخ طیفی جوهر می‌تواند برای توصیف آن و استخراج اطلاعات رنگی (به عنوان مثال تعیین اینکه آیا دو نمونه متفاوت با یک جوهر ساخته شده است) استفاده شود. بنابراین، می‌توان ترکیب جوهر را با توجه به طیف جذبی به دست آورد.

در این مقاله، توسط تصویربرداری فراطیفی نوشتارهای اسناد و مدارک در طیف مرئی مورد بررسی قرار می‌گیرد. ثبت و تجزیه و تحلیل نوشتار اسناد در صدها باند طیفی باریک انجام شده و در نتیجه، جزئیات پنهان در نوشتار سند مورد نظر بدون تماس مستقیم با آن آشکار می‌شود. مطالعات انجام شده نشان می‌دهد که تصویربرداری فراطیفی به عنوان یک ابزار غیرمخرب موثر برای افزایش قابلیت خواندن اسناد قدیمی [۴]، تغییر جوهر در مرور زمان [۵] توسعه یافته است. یکی از اولین کارها در تصویربرداری فراطیفی توسط Easton و همکاران انجام شد [۶] و یک سیستم تصویربرداری طیفی توسط Christens Barry و همکاران برای مطالعه نقشه‌های چاپ شده و اقلام فرهنگی به کار رفته است [۷].

هدف از کار انجام شده ایجاد طبقه‌بندی جوهر خودکارهای مختلف بر اساس خواص اپتیکی جوهر در محدوده طیفی ۴۰۰ تا ۸۰۰ نانومتر است. الگوریتم پیشنهادی به صورت آنی اجرا می‌شود، و احتمال تشخیص یک خط نوشته شده با جوهر هم‌رنگ و اطلاعات در مورد نوشتار یک سند جعلی را ارائه و بررسی می‌کند.

۳- شرح آزمایش‌ها و نتایج تجربی

دوربین فراطیفی مورد استفاده از سری دوربین‌های فراطیفی اسکن خطی ساخت شرکت پرتو افزار صنعت (مدل HSI-Vis-12bit) با رزولوشن طیفی دو نانومتر است که از لامپ هالوژنی به عنوان منبع نور استفاده می‌کند. این دوربین در هر لحظه از یک خط نمونه‌گیری کرده و با کنار هم قرار دادن آن‌ها یک تصویر کامل طیفی ایجاد می‌کند که به آن مکعب داده گویند. در شکل (۱) دستگاه تصویربردار فراطیفی و در شکل (۲) محل قرارگیری نمونه بر روی اسکنر در مقابل تصویربردار فراطیفی نمایش داده شده است.

برای تحلیل و بررسی یک نوشتار، در بسیاری از موارد ابتدا با آموزش نرم‌افزار یک ماتریس اطلاعات ایجاد و سپس از آن برای تشخیص رفتارهای آتی آن استفاده می‌شود. برای آموزش و مدل‌سازی می‌توان از روش‌های مختلفی مانند زنجیره‌های مارکوف، گراف‌ها و تئوری گراف‌ها، روش‌های خوشه‌بندی و روش‌های طبقه‌بندی استفاده کرد.

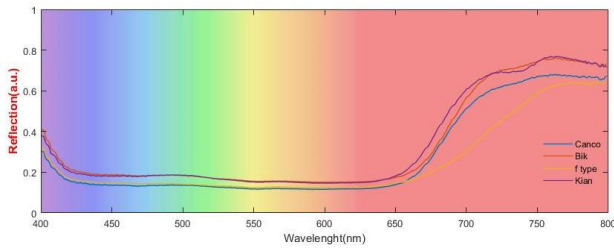
تجزیه و تحلیل جوهرها، به ویژه در بررسی اسناد، از اهمیت زیادی برخوردار است. نوع جوهر و عوامل زمانی می‌توانند شواهد مهمی در تعقیب کیفی باشند [۱]. با ایجاد تمایز بین رنگ‌های مختلف و انواع جوهر می‌توان امر اعتبارسنجی اسناد را به راحتی انجام داد. چشم انسان قادر به دیدن بخش محدودی از طیف الکترومغناطیسی است و می‌تواند بین اشیا و رنگ‌هایی که پاسخ‌های طیفی مختلفی در آن محدوده دارند، تشخیص قائل شود [۲]. با این حال چشم انسان نمی‌تواند بین جوهرهایی که از لحاظ بصری مشابه و دارای نشانه‌های منحصر به فرد طیفی هستند، تفاوت قائل شود.

به طور کلی روش‌های تجزیه و تحلیل جوهر به دو روش اصلی، تجزیه و تحلیل مخرب و غیر مخرب تقسیم می‌شود. در روش مخرب سند با روشی مبتنی بر شیمیایی مانند رنگ‌نگاری لایه نازک (TLC) [۳] توسط کارشناسان اسناد قانونی برای تشخیص عدم انطباق جوهر استفاده می‌شود. چنین روش‌هایی براساس این واقعیت است که جوهرهای مختلفی که در اسناد مورد استفاده قرار می‌گیرند ترکیب شیمیایی منحصر به فرد خود را دارند و روش متفاوتی برای واکنش با مواد مختلف بر اساس محیط واکنش دارند. با این حال، روش‌های شیمیایی به دلیل حساسیت به تغییرات دما و نیز ماهیت کیفی که دارند برای مشاهده دقیق نتایج بسیار دشوار است و همچنین در مواردی که به اندازه‌گیری مجدد و اطمینان از نتیجه نیاز است را با توجه به آسیب رساندن به اسناد از کارشناس سلب می‌کند.

برای غلبه بر چنین محدودیت‌هایی، یک روش غیرمخرب مانند تصویربرداری فراطیفی (HSI) دارای پتانسیل بیشتری است. تصویربردار فراطیفی قادر است در مقایسه با طیف‌سنجی فیبر نوری که تنها از یک نقطه از نمونه طیف می‌گیرد، در بعد فضایی هم‌زمان اطلاعات طیفی را ثبت کند. در مقایسه با روش‌های غیر مخرب همانند میکروسکوپ دو چشمی مجهز به دوربین عکاسی RGB با سه باند رنگی، طیف در کانال‌های طیفی بیشتری پوشش داده می‌شود. با پردازش تصویر فراطیفی، می‌توان از مقایسه طیف بازتابی در نقاط مختلف نمونه و تجزیه و تحلیل آن‌ها، ویژگی‌های جوهر را به دست آورد. در این مقاله با بررسی نوشتار صورت گرفته با خودکارهای تولیدی شرکت‌های مختلف کارایی دوربین فراطیفی را در تشخیص نوع آن‌ها مورد ارزیابی قرار داده‌ایم.

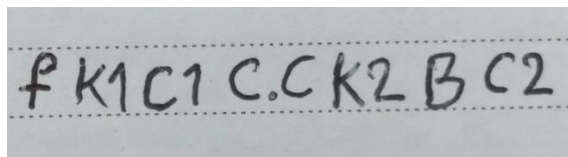
۲- تصویربرداری فراطیفی

این مقاله بر کاربرد تصویر برداری فراطیفی متمرکز است. برخلاف تصویر حاصل از دوربین های RGB، در تصویر فراطیفی از یک نمونه، هر پیکسل از تصویر شامل اطلاعات طیفی در صدها طول موج است و امکان مشاهده و تجزیه و تحلیل طیفی از نمونه‌هایی در تصاویر دو بعدی را با

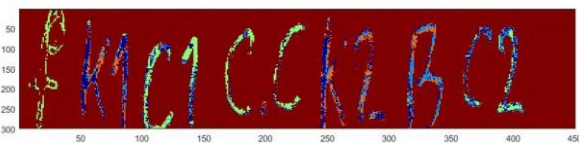


شکل (۳) : پاسخ طیفی برای ۴ مدل خودکار

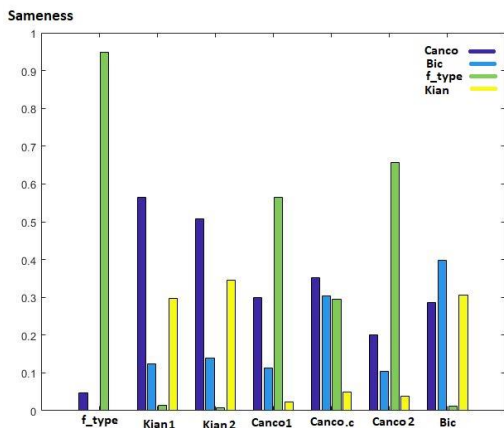
در اینجا سه نوشتار آزمایشی به عنوان نمونه برای سنجش توانایی در تفسیر تفاوت بین جوهرها مورد بررسی قرار می‌گیرد. اولین نوشتار آزمایشی از تعدادی حروف توسط هر هفت خودکار ساخته شده است (شکل (۴)) و با به کارگیری تصویربرداری فراطیفی و طبقه‌بندی SVM مورد بررسی قرار گرفته است. حرف f با خودکار f-type، حرف K1 با Kian نوع اول، حرف C1 با Cancoco نوع اول، حرف C.C با Cancoco نوع دوم، حرف K2 با Kian نوع دوم، حرف B با Bic و حرف C2 با Cancoco نوع دوم نوشته شده است. تحلیل تصویر با روش طبقه‌بندی SVM انجام شده و در شکل (۵) رنگ‌های قرمز، بنفش، آبی، سبز و زرد به ترتیب نمایانگر پس زمینه، خودکار Cancoco، خودکار Bic، خودکار f-type و Kian هستند. در شکل (۶) عملکرد تشخیص هر یک از مدل‌های خودکار توسط الگوریتم SVM نشان داده شده است.



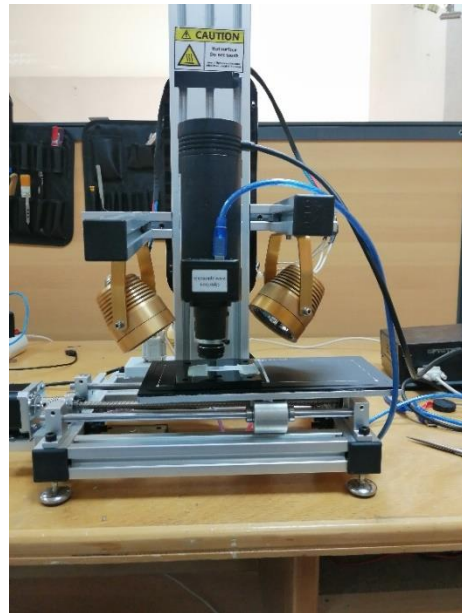
شکل (۴) : تصویر نمونه اول



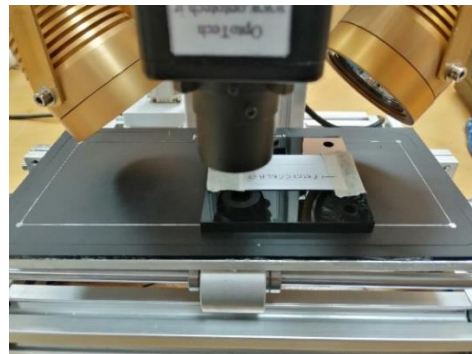
شکل (۵) : تحلیل به روش طبقه‌بندی SVM نمونه اول



شکل (۶) : مقایسه درصد مشابَهت برای چهار مدل خودکار آزمایشی



شکل (۱) : دستگاه تصویربرداری فراطیفی



شکل (۲) : محل قرارگیری نمونه مقابل تصویربرداری فراطیفی

مدل ما برای تشخیص جوهرها استفاده از طبقه‌بندی SVM است که یکی از روش‌های با عملکرد بسیار خوب برای مدل‌سازی مبتنی بر روش طبقه‌بندی است [۸]. روش طبقه‌بندی SVM که یک روش طبقه‌بندی خطی است، که داده‌های مربوط به طیف خودکارهای مختلف را از هم تفکیک کند.

برای ایجاد پایگاه داده، ما از هفت خودکار از چهار مدل مختلف با جوهر مشکی به صورت زیر استفاده کرده‌ایم: یک خودکار با نشان نامعلوم که f-type نام‌گذاری شده است، دو خودکار با نشان Kian، سه خودکار با نشان Cancoco و یک خودکار با نشان Bic. به طور کلی دو پایگاه داده متفاوت ساخته شده است، یکی برای آموزش الگوریتم و دومی برای ارزیابی. برای ایجاد پایگاه داده برای آموزش الگوریتم، خطی دلخواه بر روی کاغذ با تمام خودکارها کشیده می‌شود و تصویر فراطیفی در مکان خط کشیده شده ثبت شده و هر کدام به عنوان یک کلاس در طبقه‌بندی SVM معرفی می‌شوند.

شکل (۳) طیف بازتابی را برای ۴ مدل خودکار به کار رفته در انجام آزمایش نشان می‌دهد.

۴- نتیجه

در این مقاله بررسی و شناسایی جعل و یا تغییر در نوشتار اسناد دست-نویس ارائه شده است. روش به کار رفته روشی مبتنی بر تصویر برداری فراترکیلی و همچنین طبقه بندی SVM بود. در روش به کار رفته نمونه-های آماده شده در پایگاه داده که جوهر با رنگ یکسانی دارند به عنوان آموزش الگوریتم به کار رفته است. نتایج ارائه شده برای نمونه اول نشان می‌دهد که روش فوق توانایی تشخیص بین جوهرهای خودکارهای مختلف را دارد و نمودار توزیع برای هر خودکار در شکل (۶) نمایانگر مدل خودکار است. از بررسی نتیجه نمونه دوم استنباط می‌شود که امکان تشخیص خط خوردگی و مخدوش بودن نوشتار اسناد امکان پذیر است. تحلیل نمونه سوم نشان می‌دهد که روش به کار برده قابلیت آن را دارد که اگر در اسنادی مانند چک‌های بانکی جعل صورت گیرد، آن را تشخیص دهد.

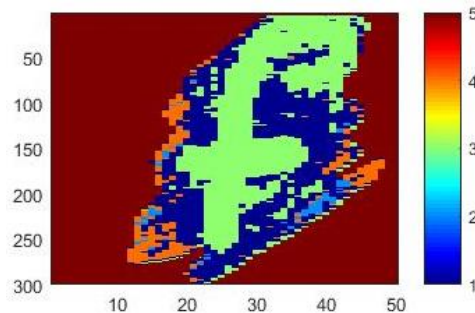
مراجع

- [1] Morales, Aythami, Miguel A. Ferrer, Moises Diaz-Cabrera, Cristina Carmona, and Gordon L. Thomas. "The use of hyperspectral analysis for ink identification in handwritten documents." In 2014 International Carnahan Conference on Security Technology (ICCST), pp. 1-5. IEEE, 2014.
- [2] Land, Edwin H., and John J. McCann. "Lightness and retinex theory." *Josa* 61, no. 1 (1971): 1-11.
- [3] Aginsky, V. N. "Forensic examination of "slightly soluble" ink pigments using thin-layer chromatography." *Journal of Forensic Science* 38, no. 5 (1993): 1131-1133.
- [4] Kim, Seon Joo, Fanbo Deng, and Michael S. Brown. "Visual enhancement of old documents with hyperspectral imaging." *Pattern Recognition* 44, no. 7 (2011): 1461-1469.
- [5] Edelman, G. J., E. Gaston, T. G. Van Leeuwen, P. J. Cullen, and M. C. G. Aalders. "Hyperspectral imaging for non-contact analysis of forensic traces." *Forensic science international* 223, no. 1-3 (2012): 28-39.
- [6] Easton, Roger L., Keith T. Knox, and William A. Christens-Barry. "Multispectral imaging of the Archimedes palimpsest." In 32nd Applied Imagery Pattern Recognition Workshop, 2003. Proceedings., pp. 111-116. IEEE, 2003.
- [7] Christens-Barry, William A., Kenneth Boydston, Fenella G. France, Keith T. Knox, Roger L. Easton, and Michael B. Toth. "Camera system for multispectral imaging of documents." In *Sensors, Cameras, and Systems for Industrial/Scientific Applications X*, vol. 7249, p. 724908. International Society for Optics and Photonics, 2009.
- [8] Cristianini, Nello, and John Shawe-Taylor. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.

در نوشتار آزمایشی دوم، حرف f با خودکار مدل f-type نوشته شده و توسط خودکار دیگری خط خورده است. شکل (۷) تصویر نمونه دوم را نمایش می‌دهد. تحلیل به روش طبقه‌بندی SVM برای این نمونه در شکل (۸) نشان داده شده است که رنگ سبز حرف f نمایانگر آن است که این حرف با خودکار f-type نوشته شده است.

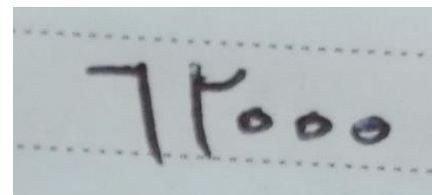


شکل (۷): تصویر نمونه دوم

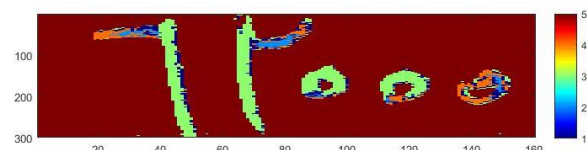


شکل (۸): تحلیل به روش طبقه‌بندی SVM نمونه دوم

در نوشتار آزمایشی سوم، عدد ۱۱۰۰ نوشته شده است و توسط خودکار هم‌رنگ دیگری به عدد ۶۲۰۰۰ تغییر داده شده است. تصویر نمونه سوم در شکل (۹) نشان داده شده است. تحلیل به روش طبقه‌بندی SVM برای نمونه سوم نیز در شکل (۱۰) نشان داده شده است. به وضوح مشخص است که این عدد توسط جوهر دیگری به ۶۲۰۰۰ تغییر داده شده است.



شکل (۹): تصویر نمونه سوم



شکل (۱۰): تحلیل به روش طبقه‌بندی SVM نمونه سوم