High-Resolution Glyph-Inspection Based Security System

Steven J. Simske, Guy Adams

HP Laboratories
HPL-2010-42

Keyword(s):
Image forensics, CMOS imaging, authentication, print parasitics, security, counterfeit detection

Abstract:

This paper describes the use of a 1:1 magnification 3.5 micron true resolution Dyson Relay lens-based 3 MPixel USB CMOS imaging device (DR CID) and software forensic image analysis system. The device enables the simultaneous capture of both intentional printing shapes and unintentional printing artifacts caused by the printing process and the interaction of the ink with the substrate on which printing occurs. The custom image analysis system written for the DR CID device allows even a single printed character to simultaneously provide fiducial marking, inspection information, authentication and forensics. We report herein on the sensitivity of the system and initial results for the reliable authentication of a printed character using DR CID hardware devices.

External Posting Date: March 21, 2010 [Fulltext] Approved for External Publication
Internal Posting Date: March 21, 2010 [Fulltext]

To be published and presented at IEEE ICASSP 2010.

HIGH-RESOLUTION GLYPH-INSPECTION BASED SECURITY SYSTEM

Steven J. Simske[1], *Guy Adams*[2]

[1]Hewlett-Packard Laboratories, 3404 E. Harmony Rd., MS 36, Fort Collins CO 80528, USA
[2]Hewlett-Packard Laboratories, Long Down Avenue, Stoke Gifford, Bristol, BS34 8QZ, UK

ABSTRACT

This paper describes the use of a 1:1 magnification 3.5 micron true resolution Dyson Relay lens-based 3 MPixel USB CMOS imaging device (DR CID) and software forensic image analysis system. The device enables the simultaneous capture of both intentional printing shapes and unintentional printing artifacts caused by the printing process and the interaction of the ink with the substrate on which printing occurs. The custom image analysis system written for the DR CID device allows even a single printed character to simultaneously provide fiducial marking, inspection information, authentication and forensics. We report herein on the sensitivity of the system and initial results for the reliable authentication of a printed character using DR CID hardware devices.

Index Terms— Image forensics, CMOS imaging, authentication, print parasitics, security, counterfeit detection

1. INTRODUCTION

Counterfeiting, warranty fraud, product tampering, smuggling, product diversion and other forms of fraud are driving the need for improved brand protection. Product security is often provided through the intentional printing of information which can later be read and decoded. Such "security printing" has additional advantages of being useful in supply chain, point-of-sale and consumer/product mobile interaction. Security printing depends on reliability of printing and reading, as well as ease of use. Readily-identifiable security deterrents, therefore, are printed so the would-be authenticators know what to read. Examples of such explicit security printing deterrents are bar codes, digital watermarks and guilloche patterns.

Other security printing technologies, such as UV and IR inks, conductive and capacitive inks, and more specialized deterrents such as adhesively attached overt deterrents, provide more limited—though valuable—roles as covert, multi-modal and/or interactive deterrents. However, these marks often require specialized readers in addition to specialized printing and/or manufacturing processes. Authentication and forensic analysis which is readily extendible to existing printing processes mitigates some of these costs and process concerns.

We herein describe a hardware/software solution that provides, simultaneously, image authentication and forensics. The hardware is described in Section 2. The software and system are described in Section 3. Results to date and a brief discussion are provided in Section 4.

2. PARASITICS AND DR CID

At the microscopic (<10 μm) level, printing on a substrate results in imperfections that can be used to uniquely identify a printed mark or glyph. Inkjet printing typically shows several classes of imperfections. Inkjet parasitics include droplet tails, which lead to narrow, curve-like ink imperfections. Extra ink deposited by the nuances of the printer and/or printing process can appear outside of the intentional boundaries of the glyph—these "satellite" ink droplets are often disattached from the printed character, glyph or other mark with which they are printed. The interaction of ink with the substrate—cardstock, paper, label stock, etc.—for printing, moreover, can lead to random wicking along paper fibers. Cellulose and other organic fibers preferentially absorb ink in the longitudinal direction (long axis) via capillary movement [1], leading to pseudopodium-like protuberances from the intended boundary, or periphery, of the printed mark. Differential absorbance of ink along the long axis of the cellulose in paper, for example, leads to parasitics such as those clearly evident in Figure 1, and often creating relatively low-ink containing "porosities" adjacent to the fiber inside of the intended periphery of the mark.

Other printing processes also exhibit variations. For instance, dry electrophotographic processes (laser) can produce multiple microscopic satellites around the periphery of a dot and liquid electrophotographic processes (e.g. HP Indigo printing) can produce small variations in dot diameter. Idiosyncrasies in printing parasitics are not limited to digital processes, as offset, gravure, flexographic, screen and other traditional printing also exhibit aberrations that can be used for reliable and robust authentication.

Effective hardware for capturing these parasitics, therefore, must be able to resolve lines with widths smaller than the smallest addressable mark a printer can form. Typically, this is measured using modulation transfer function (MTF) testing (which measures the highest frequency sine wave reliably measurable by the imaging device). In order to withstand a scan and print (copy) attack, as well as to be capable of capturing the microscopic variations with sufficient resolution for reliable authentication, the MTF of the hardware must be smaller than these printing aberrations. As the droplet tails and porosities typically range in size from 5 – 20 μm, the target MTF needs to accommodate the lower end of this range. A resolution of 5 μm equates to approximately 5000 dpi (dots/inch), which is also smaller than any printer resolution. This resolution is smaller than all but specialist scanners are capable of resolving, since mainstream scanners are typically restricted to a (true, non-interpolated) optical resolution of 1200 dpi.

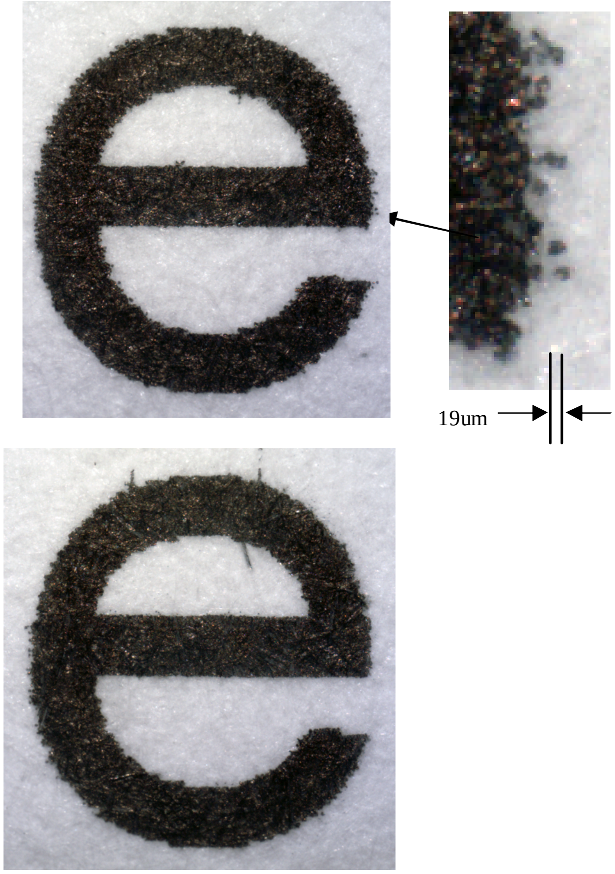

Figure 1. Two 14 pt font “e” characters, printed using thermal inkjetting and captured with the DR CID hardware. Lower images: close-ups of the tops of these characters using the forensic shape analysis software. Note the two ink-based “parasitics” on the lower image.

Accordingly, we have developed a USB-powered mini-appliance (Figure 2) that is capable of resolving spatial features of less than 5 microns with 1:1 magnification. This is accomplished using a single Dyson relay lens in series with a mirror and a low cost 3.2 μm/pixel, 3 Mpixel CMOS color image sensor. With a self-contained (white LED) illumination source, this device affords the capture of individual typed characters along with their printing parasitics.

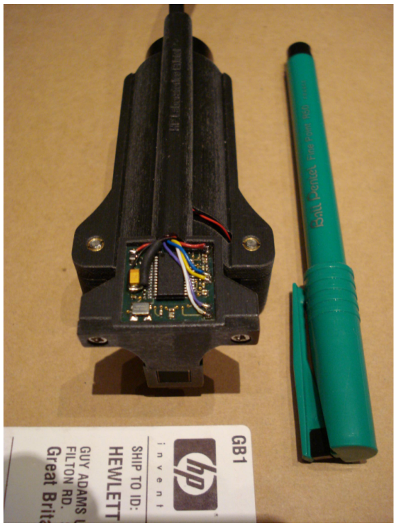

Figure 2. Dyson Relay CMOS Imaging Device (DR CID) (left) displayed alongside an office pen (right).

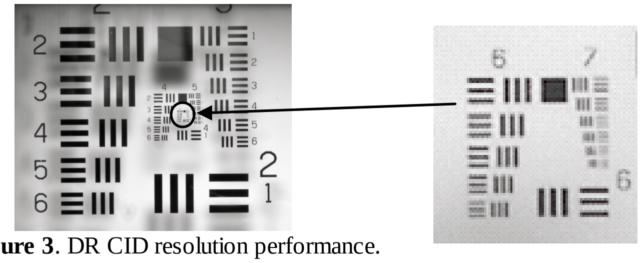

Figure 3. DR CID resolution performance.

This Dyson relay CMOS imaging device, or DR CID (named after [2]), currently provides an MTF of ≈ 3.5 microns in color. This is achieved in a handheld “contact” use model as well as providing a uniform diffuse illumination source, without unwanted reflections, through the device optics. Figure 3 shows both the field of view (FOV) and resolution of the device on the USAF 1951 resolution chart’s group 7, element 2, which is resolved at 25-30% contrast, and equates to an MTF of 288 lines/mm = 3.5 μm (7257 dpi). The inherent 1:1 magnification of DR CID also means that the variability of the scale of the images can be tightly controlled by the tolerances of the lens, further improving image analysis.

3. IMAGING SYSTEM

Images captured by the DR CID hardware are then analyzed to generate a set of printed mark features suitable for distinguishing any specific printing of a glyph or character from another. Our system accommodates non-machine readable marks (glyphs) in addition to machine readable characters and marks—essentially any mark that is intentionally printed. It can be used to analyze packaging, documents, labels or other printed items with parasitics, without the explicit need for explicit overt security printing deterrents or variable-data printing (VDP). For forensic utility, the DR CID system must be able to determine the difference between two identical glyphs printed twice, and the same glyph imaged by two different DR CID devices, or at different times or glyph orientations. Each individual character printed is therefore unique.

Custom forensic shape analysis software (FSAS) was created to provide positive identification of a unique printed item—or authentication—while providing forensic analysis (human-validation quality images). Figure 1 illustrates the forensic quality of the images captured for two different 14-point font "e" characters printed by the same printer (the first and second "e" in the name "Steve"), and Figure 4 shows the perimeters of the upper portions of these characters.

The images captured by the DR CID are analyzed by FSAS using the following steps: (1) a contrast-insensitive thresholding algorithm to binarize the image; (2) segmentation into connected components, or "regions"; (3) perimeter determination; and (4) modified perimeter shape descriptor [3] calculation.

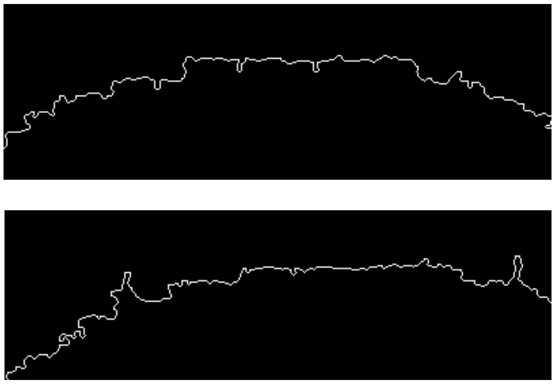

Figure 4. FSAS-determined perimeters of the upper portion of the two characters shown in Figure 1. Note the two ink-based pseudopodium-like "parasitics" on the lower images.

The thresholding algorithm assumes a bimodal distribution for the image (*I*) intensity (Int) histogram (H), similar to the Otsu thresholder [4]. To provide consistent behavior despite differences in contrast, exposure, etc., between different DR CID devices, the threshold consists of finding the 5% and 95% points in the image intensity histogram, $H\{Int_I\}$ and setting the threshold, T_I, as:

$$T_I = H\{Int_I\}|_{5\%} + 0.5*[H\{Int_I\}|_{95\%} - H\{Int_I\}|_{5\%}] \qquad \textbf{(2)}$$

After thresholding the image, the connected components, or regions, are identified and the appropriate region is selected as the glyph of interest. The perimeters are then created, as shown in Figure 4. The shape descriptors for the perimeter are next determined. The centroid of the region is computed, and the perimeter divided into sections by angle (e.g. 1° increments from 0° to 360° around the perimeter). For each angular section, the minimum radius, maximum radius, complexity (number of changes in direction of the perimeter in radial direction with the glyph centroid as the origin), shared elements (number of perimeter points in the section), uncertainty (number of perimeter line segments in the section), and neighborhood uncertainty (moving average of the uncertainty to account for minor, i.e. less than 0.5°, differences in alignment of the two images with the angular sectioning) are computed.

When two images are to be directly compared, the second image is scaled to the first image to match connected component size. This "normalization" corrects for any difference in focal length between two DR CID devices; difference in height of the DR CID devices over the glyph during image capture; and difference in size of the glyph, e.g. due to font, ink gain, etc. differences. After normalization, the images are aligned by angle (least squared error for the difference in their features) and the same set of salient parameters—maximum and minimum radius, complexity, shared elements, uncertainty and neighborhood uncertainty—is computed. We denote these Max-R, Min-R, Cmplx, ShElem, Uncert, and Neigh-Uncert, respectively.

Another set of image data is then computed for these "normalized" images. For four of the features—maximum and minimum radius, complexity and neighborhood uncertainty—the differences between the two images in all the angular sections are computed. Mean differences, denoted ΔMax-R, ΔMin-R, ΔCmplx, ΔShElem, and ΔNeigh-Uncert are used to compare goodness of fit. Absolute differences are computed, and are used to identify possible "satellites" in one image compared to the other.

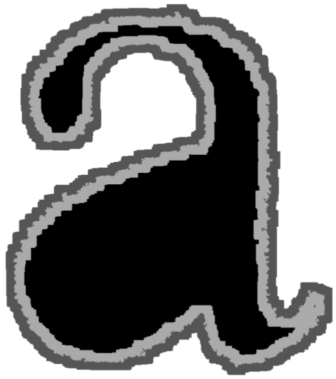

Figure 5. Character "a" with salient perimeter search areas indicated: porosities in the lightest (middle) zone, just interior to the glyph perimeter; and satellites in the outermost zone, just exterior to the glyph perimeter.

Additionally, the Uncert and Neigh-Uncert features were "thresholded" (only those >= 1.0 standard deviations above the mean for the angular sections are retained) and compared as features T-Uncert and T-Neigh-Uncert.

Satellites and porosities are then identified from the glyph images after "search zones" are identified as shown in Figure 5. Search zones extend inward or outward from the periphery, and the zone width is determined as a fraction (typically 20%) of the mean width of the glyph. Satellites are determined from the previously described connected component map as regions of non-trivial size lying in the outermost zone of Figure 4. Porosities are "inverted" connected components created inside the glyph (i.e. they are regions whose pixels are lighter than threshold intensity). In comparing two images, locations of absolute differences in ΔMax-R, ΔMin-R, ΔCmplx, ΔShElem, and ΔNeigh-Uncert are noted and added to the potential "satellite" list. The Hamming distances between two images are computed based on location differences in satellites and porosities. These are designated HD-Sat and HD-Por, respectively.

4. RESULTS AND CONCLUSIONS

In order to test the relative value of each of the features for identifying authentic glyphs, we performed the following experiments. A DR CID device was used to capture 10 different 6 pt "a" characters on an HP K5400 thermal inkjet printer. Each unique character was captured four times, with differing placement, rotation, white balance gain and focus, corresponding to $C_{4,2}$=6 comparisons the "exact" characters for each of 10 unique characters. Among these 40 images, then, there are $C_{40,2}$=780 comparisons (60 of which compare the exact same printed character and 720 of which compare different printed characters).

Feature Name	Matched (60)		Unmatched (720)		n Stds Apart
	Mean	Std	Mean	Std	
ShElem	0.573	0.020	0.793	0.021	10.5
Neigh-Uncert	0.971	0.006	0.881	0.016	5.7
T-Uncert	0.917	0.024	0.809	0.021	5.2
ΔMax-R	0.073	0.025	0.180	0.020	4.3
HD-Sat	10.6	5.0	29.1	4.2	3.7

Table 1. Imaging features and their ability to distinguish (n Stds apart, using larger of the two Stds) between matched and unmatched "a" glyphs. All features shown are statistically significantly different at $p<10^{-6}$, modified t-test, df = 8.

These experiments were repeated for other character sets, with similar results. For the widely different "k", "s" and "l" characters, for example, ShElem was 9.9, 9.8 and 9.4 Stds apart when comparing matched and unmatched characters. These experiments demonstrate DR CID usage for forensic image analysis: the captured image is compared either to the original image (i.e. if associated with another serial number) or else compared to a set of images if individual item serialization is not possible. The features generally providing the greatest statistical power for identifying the matching images are given in Table 1, and the ShElem feature results, also for "a", are given in Figure 6.

The FSAS software resolves differences between camera-to-camera variance and glyph-to-glyph (or character-to-character) variance. Our ongoing research focuses on determining the best overall set of features to use; that is, the set which provides the highest classification accuracy. This set will likely differ based on the end goal: e.g., glyph authentication as shown here, printer forensics and ballistics [5,6,7], or substrate forensics [8].

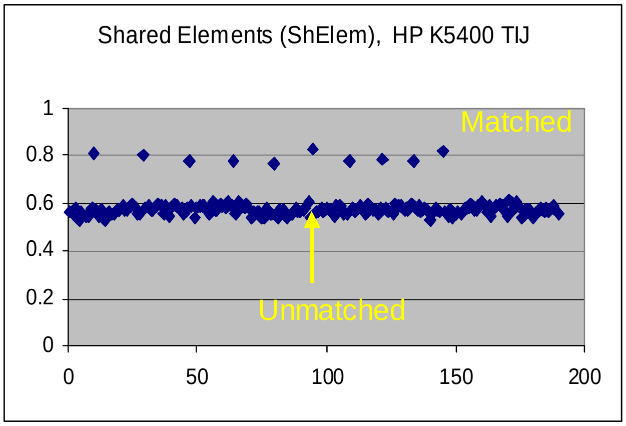

Figure 6. Sample matching data for the feature "shared elements", or ShElem. When a specific "a" character captured is compared with an image of the "a" captured earlier, the match is 0.792 ± 0.021 ("Matched"). When it is compared to an image of one of the other 9 "a" characters, the match is 0.573 ± 0.020 ("Unmatched"). The two populations are 11.2 standard deviations apart.

Moreover, the DR CID+FSAS approach can be tied directly to image quality and inspection, since it is well-suited to individual text characters. In fact, using a spot color with an overprinted character simultaneously provides text print quality assessment and a fiducial mark for the rest of the inspection/forensic process.

6. REFERENCES

[1] C. Skaar, *Wood Water Relations*, Springer-Verlag, NewYork, 283 pp., 1988.

[2] CID, Criminal Investigation Department in the UK, see http://en.wikipedia.org/wiki/Criminal_Investigation_Department

[3] S. Belongie, J. Malik, and J. Puzicha, "Shape Matching and Object Recognition Using Shape Context", *IEEE Trans. PAMI* April 2002, 24(4), pp. 509-522.

[4] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Trans. Syst. Man Cybern.*, Vol. SMC-9, No. 1, 62-66, 1979.

[5] W. Deng, Q. Chen, F. Yuan, and Y. Yan, "Printer Identification Based on Distance Transform," in *Proc. 1st Intl Workshop Intell. Networks Intell. Systems*, pp. 565–568, Nov. 2008.

[6] N. Khanna, A.K. Mikkilineni, A.F. Martone, G.N. Ali, G.T.-C. Chiu, J.P. Allebach, and E.J. Delp, "A Survey of Forensic Characterization Methods for Physical Devices," *Digital Investigations*, vol. 3, pp. 17–28, 2006.

[7] E. Kee and H. Farid, "Printer Profiling for Forensics and Ballistics," *Proc. ACM MM&SEC'08*, 2008, pp. 3-10.

[8] W. Clarkson, T. Wyrich, A. Finkelstein, N. Heninger, J.A. Halderman, and E.W. Felten, "Fingerprinting Blank Paper Using Commodity Scanners", in press for Proc. IEEE Symp. Security Privacy, May 2009, http://citp.princeton.edu/pub/paper09oak.pdf.