# Secure color. Advanced color techniques for analysis and security.

Luis Borruel[i] and Miguel Ambrona
Fábrica Nacional de Moneda y Timbre. Jorge Juan 106. Madrid 28009. Spain

## ABSTRACT

Color is present in all security documents. Most of the times we use our own eyes to assess the color characteristics of the document, but our visual skills to concentrate on the color accuracy are quite limited. Automatic systems have not provided either a procedure to color-assess a complex document in a simple way. In this work we present how to create a simple, fast, meaningful and accurate color fingerprint of a document by using three bivariate histograms. This procedure uses simple desktop equipment and has a wide range of applications, for example: color quality control of production, counterfeit detection or soiling assessment. Being this procedure especially sensitive to go beyond what our eyes can see, documents with especially crafted color content could reveal unique fingerprints difficult to replicate. This new generation of security features is called Hidden Histograms. Its strengths, difficulties and possibilities to embed digital information are discussed.

**Keywords:** Color management, Color Calibration, Bivariate histograms, Soiling, Counterfeit Detection, Security Feature

## 1. INTRODUCTION

Color is behind the main interaction of security printers and end users. Despite color is an everyday experience, color characterization of documents (whether or not they are security documents) lacks of powerful but simple tools to ease document study from a strict colorimetric point of view.

Probably colorimetric characterization brings to memory spectrophotometers, laboratory equipment or control strips. None of them deal with easy to understand concepts and, more important, none of them make it possible to characterize a document, as a whole, like some sort of *color fingerprint*. We can measure spot colors, wedges and so forth, but there is no way (except may be the good old subjective comparison with the reference sheet) to judge and numerically evaluate the complete color document.

This work was born with the main idea of developing a visualization tool to characterize, from a colorimetric point of view, (security) documents subject to:

- Generate discernible information of the complete document, either by people or machines.
- Require simple desktop equipment.

The potential benefits of having such tool are paramount, such us full-document color quality control, document authentication or the development of *ad-hoc* new security features making use normal printing equipment.

## 2. REVIEW OF CURRENT TOOLS

Color is mathematically represented by color coordinates linked to a certain color space. When a color space is defined for a certain set of colors, either the complete set of discernible colors or any smaller subset, we link that human experience to a certain mathematical domain in $\mathbb{R}^n$. Therefore we can distinguish *color* as a human experience, and `color` as a mathematical representation of a human experience. We shall try to distinguish them by the different face convention when the term can be confused. Each `color` within a certain color space is determined by n coordinates. Example of well-known color spaces are RGB (with Red, Green and Blue coordinates) which is used by most capturing devices and displays, and CMYK (with Cyan, Magenta, Yellow and black components), typically used by printing devices. Both of them are known as device-dependent color spaces as they do not unambiguously define a *color*[1]. In spite of that, each coordinate, or "primary", is meaningful, i.e. Red or Magenta makes sense to all of us as *colors*.

[i] Luis Borruel: Luis.borruel@fnmt.es. Jorge Juan 106. Madrid 28009. Spain. www.fnmt.es

Other color spaces, such as CIE XYZ and CIE L*a*b* are device-independent or, in other words, can unambiguously define a *color*. Unfortunately X or a* are no longer identified with a direct human experience. L*a*b* was defined by the *Commission Internationale de l'Eclairage* (CIE) in 1976 as a derived space from XYZ that pursued higher linear behavior from a perceptual point of view, i.e. that equal numerical increments would be linked to equal changes in the *color* experience . L*a*b* coordinates locate a point in a 3D Cartesian space. If the same space is expressed in cylindrical coordinates, we have the L*c*h* color space. The main advantage is that the three coordinates are more meaningful. L* is the same as in L*a*b*, but now c* (chroma) —the radius in the cylindrical coordinates— accounts for the saturation and h* (hue) —the angle in the cylindrical coordinates— accounts for the color shade going through all the rainbow colors.

L*a*b* color space is one of the key elements of Color Management, which pursues having consistent color reproduction in a complete workflow, starting with image capture, edition & design and finally printing. It is a very extensive topic that would require many pages to be thoroughly described; there are numerous references elsewhere[1] for that purpose. It makes use of color profiles, or calibration files for every input/output device —like monitors, scanners or printers— and the so-called rendering intents for color transformation among different color spaces.

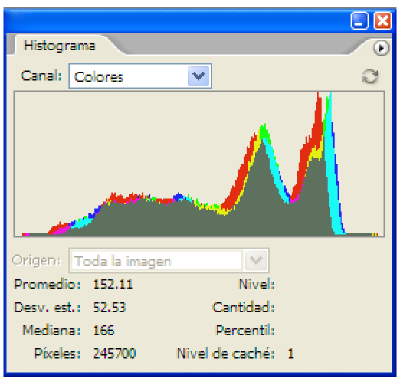



Figure 1. Capture of a 1D Histogram of RGB values

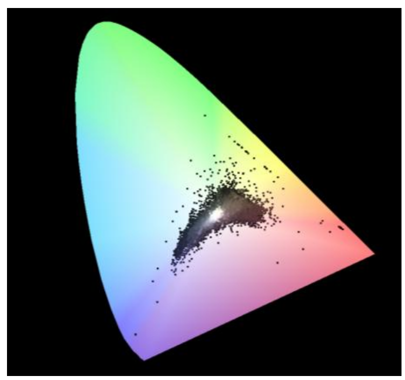


Figure 2. CIE 1981 XY projection of image pixels

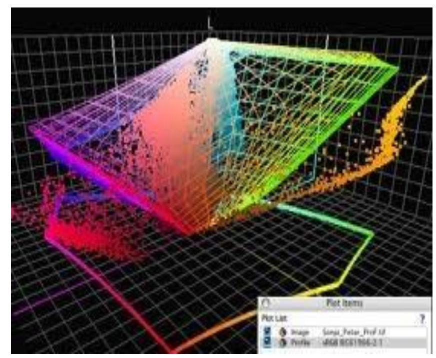


Figure 3. View in 3D of the CIE Lab image pixels (Courtesy of www.colorwiki.com)

So far there have been available a limited number of tools to perform some color analysis on images. Probably the best known tool is a 1D histogram (see Figure 1). They are available in Photoshop or any other image editing software. It gives some feedback on the shadows-highlights distribution of the image with a simple pattern. In spite of that, it barely gives any information about the color characteristics, dominant colors, saturation, etc.

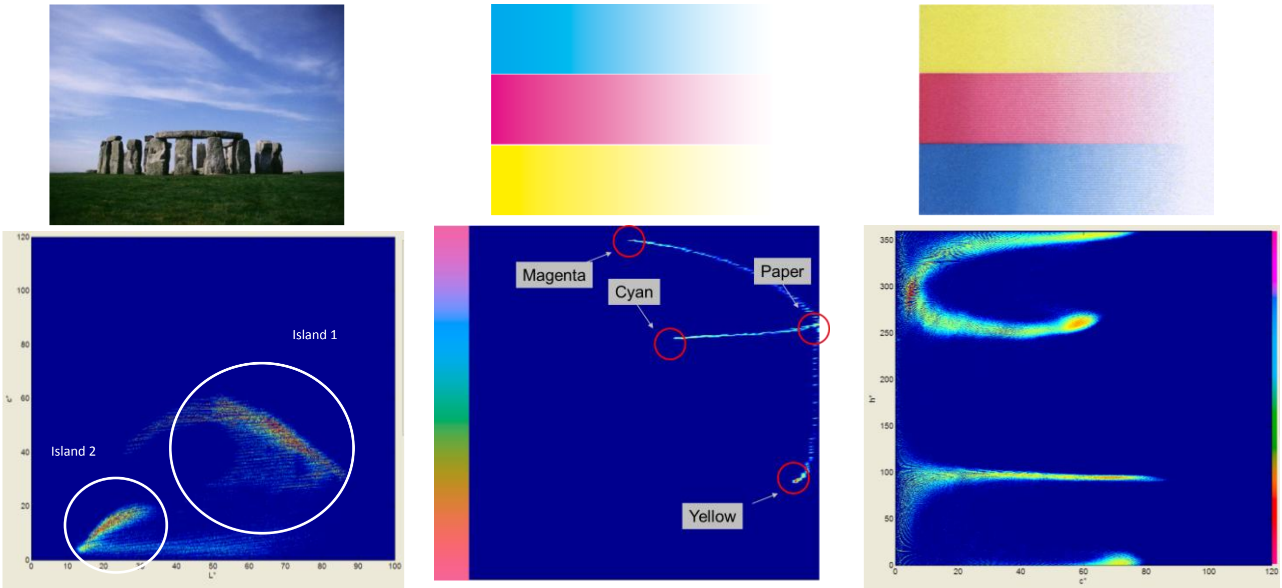
More specific tools like *Colorthink*[2] or *Matlab*[3] perform custom conversion among color spaces and representation in 3D view or in 2D projections. Conversion from or to device-independent color spaces is usually supported by the color profile associated to the image. Pixel projection onto the CIE XYZ space (see Figure 2) provides, unlike the 1D histogram, accurate information about the colors of the image, but XYZ coordinates are difficult to interpret as they are related with physical models of the light receptors in the eye.

Different drawbacks can be stated for a 3D view of the L*a*b*/L*c*h* color space (Figure 3). Indeed we have access to the full range of color information, accurate information; however, a 3D space is always difficult to represent in our communication media (usually 2D). In addition, even if this cloud of points could be easily visualized, is not an easy task to compare things, neither by people nor by machine algorithms.

## 3. BIVARIATE COLOR HISTOGRAMS.

The cornerstone of this work lies around a set of 3 bivariate **Col**or **His**tograms (ColHis): *Lc*, *Lh* and *ch*. Each of these histograms provides a wealth of information about the image in a clear and recognizable way, as it will be shown.

Unlike 1D histograms, which are commonplace in image edition, 2D histograms or bivariate histograms[4] are quite seldom in the color literature. Searching for use of bivariate histograms in color images barely reports a single reference in the past 15 years (on an image segmentation application)[5]. In a bivariate histogram the frequency, or repetition rate, of




*Lc* histogram *Lh* histogram *ch* histogram

Figure 4. Example of three target images (upper row) and one of the three color histograms in each case

pixels is represented as a function of two variables. Taking into account that these two variables will go in the x and y axis, the frequency/repetition needs a third coordinate. In order to keep graphs flat, the repetition is associated to a colormap, where bluish (i.e. cold) colors represent low repetition and reddish (i.e. warm) colors represent frequent repetition.

The Figure 4 is meant to be illustrative of the qualitative information than can be obtained by each of the three ColHis. The left column deals with the first target image, which is part of the Windows XP wallpaper collection (Stonhenge.jpg). For the analysis it has been assumed that the image is associated to the *sRGB* color space (the default color space in the Windows and the web). Below the image it is shown the first ColHis shown so far in the report. The horizontal axis corresponds to Luminance (L*), and the vertical axis to the Chroma (c*) value. The first noticeable thing is that most of the graphs show a blue background. That blue region means values of L* and c* not represented by any pixel in the image (Stonehenge); in other words, no occurrence of those *Lc* combination are being produced —it must be kept in mind that the cold colors identify low occurrences and hot colors high occurrences—. The regions with colors different from dark blue identify values with at least some pixels having the corresponding *Lc* pair. A general tip observing *Lc* ColHis is that the left part of the histogram corresponds to pixels in the shadows region (low L*), whereas the right part corresponds to the highlights (high L*). In consequence, the bottom of the graph (low c*) correspond to the gray-neutral parts of the image, whereas getting close to the upper part of the graph matches highly saturated (high c*) parts of the image. Two "islands" have been marked onto the figure, what do they mean? Obviously they have to correspond to two predominant color shades in the Stonehenge image, namely: green (from the grass) and blue (from the sky). As the green shades are darker than the blue ones, Island 2 will match the grass area and Island 1 will match the sky area. In addition to the difference in luminance (x axis), blue colors have higher saturation. A cyan band in the lower part of the histogram, having very low c* values, i.e. colors with low saturation, matches the stone grayish parts of the picture. After some training checking ColHis, it becomes easier to interpret the results.

In the second column, the target image is a very simple synthetic Cyan-Magenta-Yellow grading to white. In an *Lh* histogram the horizontal axis is also corresponding to Luminance (L*), but the vertical axis carries the Hue (h*) value. Contrary to L* where everybody understands that L* = 0 is black and L* = 100 is white; learning that h* = 50 corresponds to red, or h* = 250 to cyan is difficult to remember. That is the reason of the rainbow-like band in the right side of the histogram. It can be observed that the ColHis is made of three "lines" joining the white point (paper) with the three solid primary colors. That is what should be expected from a linear grading. It is important to remark that the transition does not follow, in general, a straight line. A very important fact pointed out by this histogram is that the white paper has a certain hue. That property is intrinsic of the paper fabrication.

The table concludes with a real grading (the one shown in the right column), printed with a laser CMYK printer and later on scanned. It will be used to show the remaining histogram: the *ch* histogram. In a *ch* histogram the horizontal axis corresponds to the c* coordinate (saturation) whereas the vertical axis does with the h* (hue). Note that being h* an angle, there is continuity between h = 0º and h = 360º. It is clearly seen that there are three horizontal bands showing pixels with a relatively constant hue and decreasing (left-wise) levels of saturation. They correspond with the three primaries from the printer (cyan, magenta and yellow) being progressively mixed with the color of the paper. However, the three bands have different properties, the highest saturation (c*) is achieved for the yellow color, next is magenta and last one cyan. It is also noted that the purity of the yellow is superior, observing the little expansion across h* values. Those are properties directly related to the pigment base of the tonner. To what extent for each h* value a histogram can spread to the right side of the graph (higher c*), deeply depends on the inks and printing process properties. Needless to say that it also depends on the image contents, but it is ultimately limited by the pigments involved and how effectively are attached to the substrate.

Finally this qualitative analysis tackles a security document**,** a real scanned banknote; in Figure 5 we can appreciate the *Lh* histogram of the front side of the 1 Yuan banknote (300 dpi scan 24-bits depth)

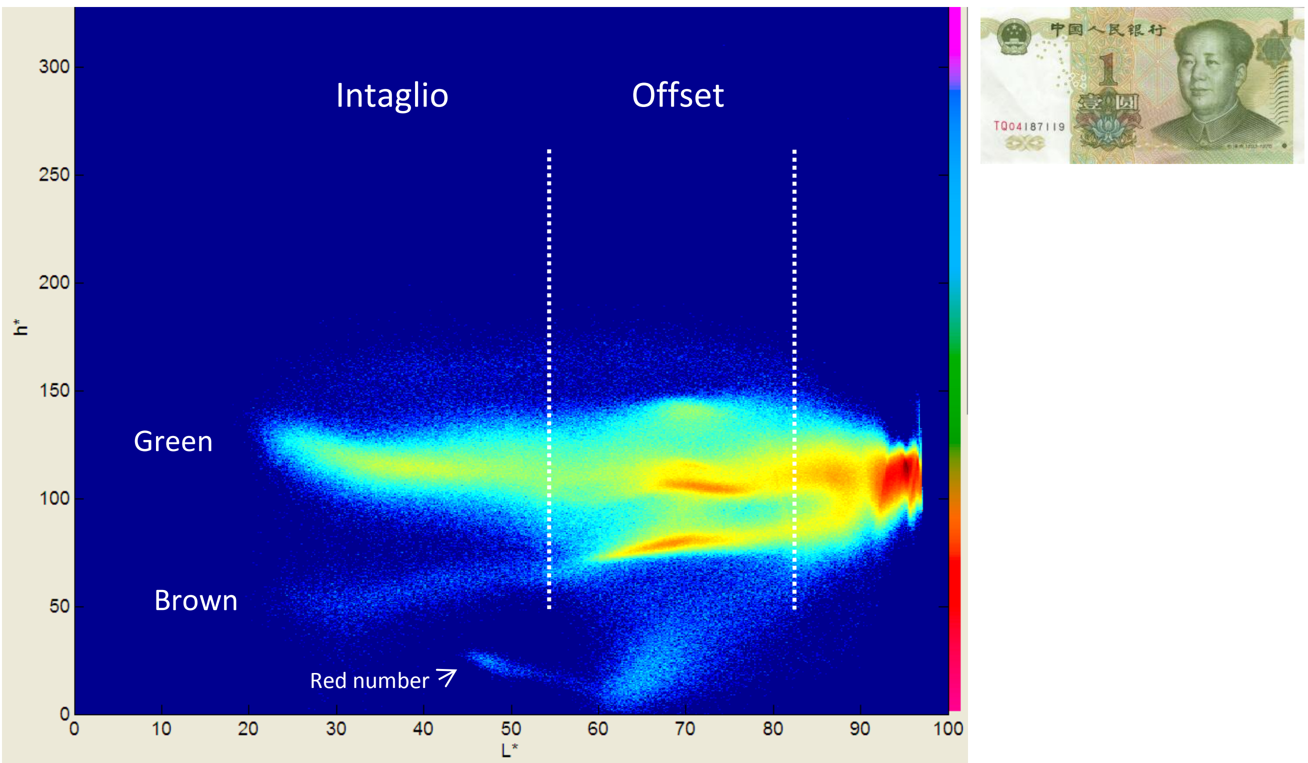



Figure 5. *Lh* histogram of a 1 Yuan banknote. Annotations to identify the different areas

The banknote has a predominant green-orange component, as can be confirmed in the histogram, only the reddish letterpress in the numbering is slightly separated from the horizontal color bands in Figure 5. It can be appreciated two groups of inks: intaglio (with lower luminance values) and offset (with luminance close to 70). The high concentration at L=95 corresponds to the white paper. The reason for having soft transitions between the different islands is mainly due to the resolution. Security documents are mainly created with very fine lines, instead of solid large areas; those lines are discriminated in a digital capture only at high resolutions. When the resolution of the capture drops, the offset backgrounds start to blend with the substrate color. The effect of this blend in the histogram is appreciated as blurring.

So far visual evaluation of the ColHis provides a lot of information at a glance, and makes relatively easy the comparison between two documents. For machine use and, in general, for providing a qualitative figure for the comparison, it is good to define **F**igures **o**f **M**erit (FoM), or parameters that account for certain properties of the histograms. The four properties observed that should have a number associated would be:

- Spreading. ColHis can be sharp or blurry, and that can be characterized with a number proportional to the surface covered by the ColHis.
- Color of paper or predominant color. Any image has a certain color, or shade, that shows a higher number of pixels, something like the average color of the sample.
- Distribution of colors. Taking as a reference this predominant color, how the rest of the colors distribute in the surroundings of it.
- One type of histograms: *ch*, can give information about the physical properties of the pigments/dyes used in the printing process. The highest saturation values (c*) for each color shade (h*) is probably the most crucial colorimetric property of the image. In the *ch* histogram that information is provided by the right-most parts of the *ch* histogram.

According to those needs, certain FoM were identified and implemented (Figure 6 shows a graphical representation of each FoM, providing visual feedback of what features characterize):

- Area: a certain threshold is set and the area inside is calculated. It is equivalent to iso-level terrain lines.
- Bogus **c**enter of **m**ass (CoM): formally, the center of mass —where mass in our case is substituted by frequency of repetition— defined as $\vec{X} = \sum_i m_i \vec{x_i}$ , does not show a good correlation with the visual appearance of the color histogram, so it was modified to lower the weight of the mass parameter.
- Polar diagram: Centered in the (bogus) CoM this parameter, a polar diagram gives feedback on the distribution of the pixel frequencies. It bares resemblance with a moment of inertia angle-diagram or the radiation diagram of an antenna.
- **E**nvelope **d**iagram (ED): contour of the maximum saturation per hue value.

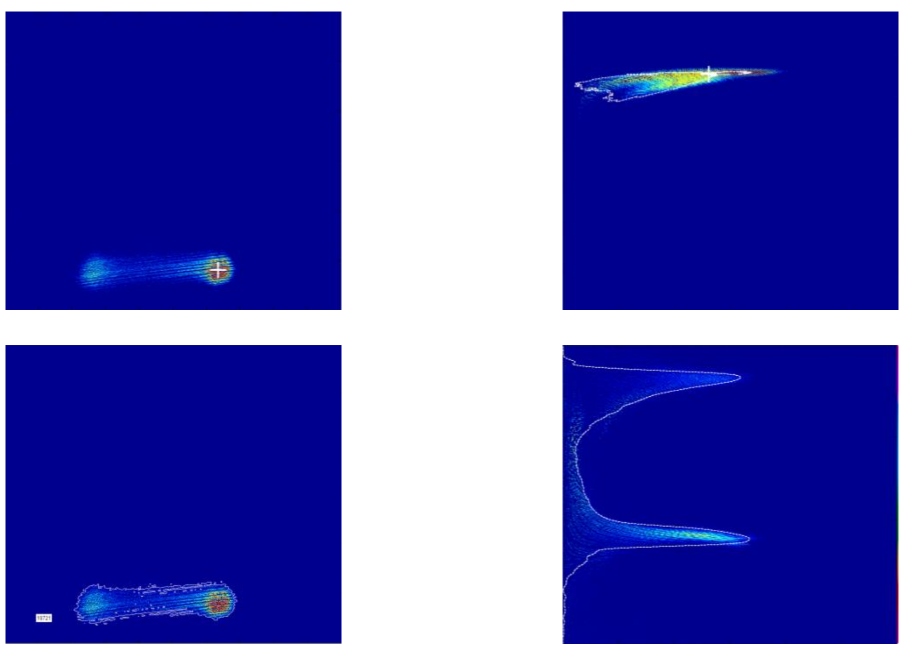


Figure 6. Graphical representation of the 4 FoM: a) upper left: center of mass; b) upper-right: polar diagram; c) lower-left: area and d) lower-right: contour plot or envelope diagram.

These FoM will make possible to compare documents, for example, taking the CoMs of two documents, we could calculate the Euclidian distance of two CoMs and use it as a similarity parameter.

In addition to the availability of FoM, the histograms can be filtered-out using ranges for each $L$, $c$ and $h$ variable. For example, should we want to filter out from the analysis the influence of paper, we can remove higher than $L > 90$ values.

## 4. APPLICATIONS IN A PRODUCTION ENVIRONMENT

In Graphic Arts, color control of the final product is currently restricted, if any, to the verification of color strips printed at the sides of the sheet. This assumes that having good reproduction of the auxiliary patches will correlate with good consistency of the final product color. Notwithstanding this reasonable assumption, this methodology does not conceal the fact that there would be no other way to color control without single patches. To the best of our knowledge there is no way to monitor color deviation of a color full-document, for instance, a picture. Such method should face a difficult trade-off: accounting for the tolerable variations of the printing process, paper and noise captured, while pinpointing severe drifts of the complete set.

ColHis is a reasonable methodology to first-ever achieve color monitoring of a full color document. The secret relies in the intrinsic statistical character of ColHis, where we can see at a glance the trends of the color contents, the average values and the dispersion around those values. The logarithmic representation makes it possible to locate all the color features, and if needed, isolate some areas for better discrimination.

Provided enough capture resolution (color discrimination improves with resolution), the same methodology could be applied to detect counterfeits. The reproduction process will include detectable changes in the ColHis, even changes unnoticed to the naked eye, which is not especially trained to detect saturation accurately. We proved, in a pilot study of euro banknotes (not published so far), that the differences of the best counterfeits are far larger than production changes among different printers or years. If the reproductions are carried out by screening (digital printing, for instance) instead of spot colors, the differences are even more obvious. A refined definition of the FoM could be even useful to sort counterfeit classes.

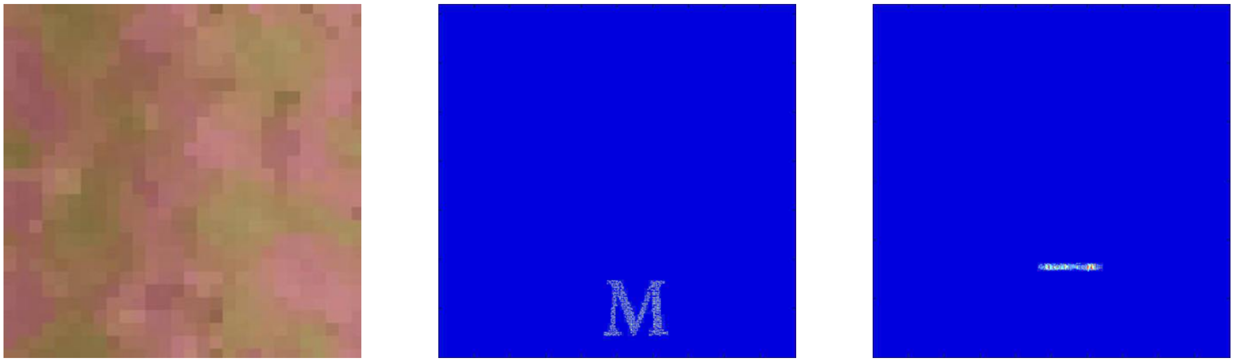
A well-known current problem in the cash-recycling scenario is the correct evaluation of fit and unfit notes returning from the commercial circuit. Soiling is the main reason for taking banknotes out of the circulation (i.e. classified as unfit). Recirculation of unfit notes is undesirable, but the destruction of fit notes is even worse, with a high monetary and environmental cost. So far the evaluation strategies have been somehow bipolar, using very simple sensors checking simple parameters, or performing very complex analysis. Previous studies have pointed out that soil have a strong yellowish component, but there is not a good established framework to evaluate soiling, either at the watermark (i.e. printing-free) area or within the complete note surface. In the latter case, the watermark area is a good representative of the global dirt in the banknote, but it obviously will miss local stains in other parts of the banknote. This scenario seems like a good testing field for ColHis. Grease, dirt, folds and stains are expected to leave detectable footprints in the color characteristics of the print-free area and the global note.

## 5. NEW SECURITY FEATURES: HIDDEN HISTOGRAMS

Once it has been shown the powerful analysis that can be extracted from the 3 ColHis of an image, further extended with the definition of FoM, a natural consequence is the question: is it possible to create specific security features to be analyzed using ColHis?

At early stages of this development it was clear that the ColHis themselves, the plots, were also images; to some extend the original image was transformed to a different domain, therefore would it be possible to create “singular” or “unique” ColHis? Hidden histograms (HH) are the answer to that question.

A hidden histogram is a complex concept at first. Although is easy to understand what they are, it can be cumbersome to understand the rationale behind them. For the sake of clarity, suppose we could generate a digital image (Figure 7a) with a distribution of colors showing no apparent pattern (although it is not a restriction: indeed there could be a pattern), it could be an RGB image associated with some well-known color space profile, like *sRGB* or *AdobeRGB*. If the ColHis of that image is calculated and shows a meaningful pattern or shape, we would have a HH. The pattern could show up in only one of the ColHis (*Lc*, *Lh* or *ch*), or it could show up in several histograms simultaneously. In this figure the *Lh* histogram shows part of the FNMT-RCM logo and the *Lc* histogram, simultaneously, a single line.



RGB image (a) *Lh* histogram (b) *Lc* histogram (c)

Figure 7. Color image with no apparent pattern. Two color histograms of the image

Before entering into the details of generating HHs, it is useful to explain the reason why this test image from Figure 7(a) has shown up a HH. The pattern popping up in Figure 7b, for example, is just a consequence of the color contents of the image (Figure 7a). The pattern we see in the HH is associated with a set of colors (in Lch coordinates) that are indeed present in the digital image. As they are present, they have a non-zero frequency in the histogram. The blue background in Figure 7b is associated with colors (Lch coordinates) that are NOT present in the image, that there are no occurrences of those values. Therefore, the RGB values of Figure 7a are not really random, they are a carefully chosen population that in RGB does not make any sense, but whose frequency distributions in Lch follow very clear patterns. It is important to remark that this initial RGB image would give rise to a HH, but any arbitrary rearrangement of the pixel locations, would give rise to the same HH (at least in this theoretical examples we are dealing with). The reason is that the HH shows up because of the pixel contents, not because the pixel spatial distribution.

Generation of HH (*Lh* histogram for example) requires the following input from the end user:

- Definition of a binary (1 bit) mask containing the hidden pattern.
- Domain/region in the *Lh* space in which the pattern will be mapped.
- Color space in which the digital image will be created (for example, *sRGB*, some printer's RGB/CMYK color space, etc…)

It would be possible, in principle, to define the pattern in gray scale, but this rather complex approach has not been developed, and only binary masks were used. The first step would be to collect a list all the $L^*$ and $h^*$ values lying within the mask in the chosen domain, i.e. the maximum and minimum values. Needless to say it must be a finite number of colors. After collecting the values for L* and h*, let assume we find N different pairs, we have to revert the color coordinates to the destination color space, using for instance a color profile (.icc).

Lch → Lab → destination color space (RGB for example)

But before performing that color conversion we need to assign some value for the remaining Lch coordinate, in this case the *c** coordinate. The simplest approach would be to set a constant value that satisfies that every triplet $(Lch)_{1..N}$ is part of the destination color space. It would not have much sense to locate our HH in a region outside the range of the destination color space; otherwise the Lab → destination conversion could not be accurately carried out.

This simplest case in which the first two coordinates are chosen depending of a certain mask and the third set to a fixed value is what we generally call <u>single HH</u>. In such HH, one ColHis will show the pattern and the other two ColHis will show two straight lines.

A refinement of the process described above would lead us to higher sophistication, where the same pattern is simultaneously shown in two different ColHis (*Lc* + *Lh* for example). This is called a <u>double simple HH</u>. Further refinement makes it possible to show two different patterns. There are geometrical constraints that force some compatibility between those two patterns, as they cannot have arbitrary shapes. In particular they must have identical projection on the common axis. This restriction is not particularly tight and makes possible a huge number of compatible patterns. In this case it is called a <u>double complex HH</u> (see Figure 8).

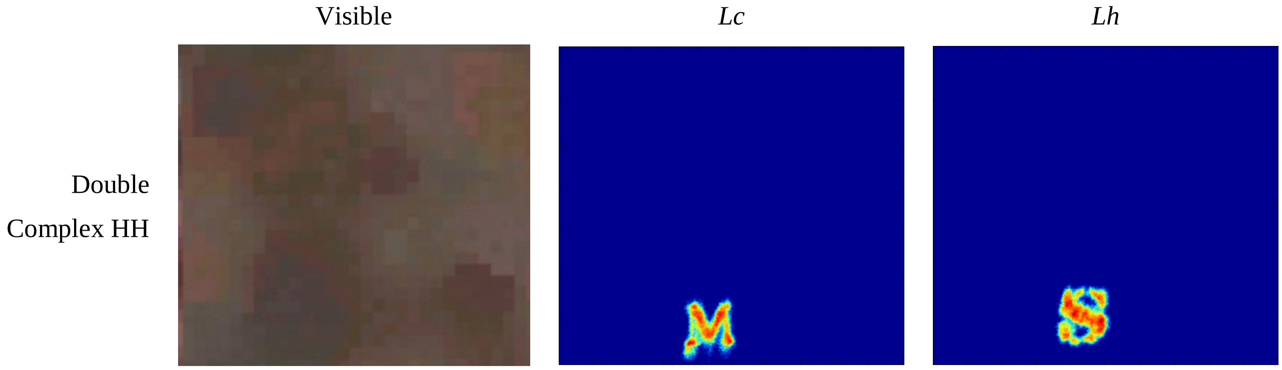



Figure 8. Digital capture image of a double complex ColHis, showing an "M" symbol in the *Lc* histogram and an "S" symbol in the *Lh*. The remaining histogram (*ch*) is not shown because it does not show any significant pattern.

So far we have only created a digital file containing a HH, and although it poses a number of difficulties, it cannot be readily used as a security feature. This file needs to be printed, and it has to be accurately printed. The main physical requirement to print HH is to have available a very carefully calibrated printing device. All the tests carried out so far have been done using inkjet printers (firstly a photo-quality desktop printer and later on a large-format professional printer) and the first calibration approach involves using ICC color profiles.

ICC calibration profiles were created with the main goal of serving a color management system whose beneficiaries are people. They were not meant to linearly reproduce the L*c*h* color space, in consequence ICC profiles might not fully exploit the accuracy delivered by a printer. Custom calibration procedures could improve the performance accuracy of the printing process.

Calibration is only possible if the device has enough repeatability. Our studies have found that a repetition figure of $\Delta E \leq 1$[ii] is enough to print reliably this security feature. In principle every printing device with significant repeatability and potential to smoothly cover a color space could be suitable to print HHs. The later requirement practically forces to use screening strategies through a printer driver or a raster image processor. Size of printing is another important factor. It has to be a trade-off between reducing the overall size of the security feature (which is always desirable to reduce the visual impact and the integration in previous designs) and keeping accuracy in the subtractive color synthesis. Every tile in a HH has to be, at least, large enough to guarantee a stable average color. Should the tile is further reduced the synthesis begins to fail and the color of the paper or the color primaries become significant. A simple example: if we zoom in a half-tone image of a green color printed with CMYK we get to a point in which green is no longer seen and we start to perceive white paper, cyan and yellow. The minimum tile size is eventually limited by the printing device and the substrate chosen. In our development this tile size limit was about 200 µm and the smallest HHs which we have successfully created are 1.5 cm width.

Once the printing stage has been solved, the next essential step is the verification process which implies a digital capture of the document. Examples of capture devices would be scanners and digital cameras. The capture process poses a number of challenges to properly recover the color information: the illumination homogeneity, the sensor imperfections (such as noise or non-uniformities) and the Point Spread Function (PSF) of the optical system. Out of the three, the last one is the most detrimental and difficult to correct.

The most reliable capture device, mostly for the controlled illumination conditions, is a desktop scanner. In our case the scanner used was an EPSON Perfection 4990. A scanner provides an environment with reasonably homogenous and repeatable illumination. This strong illumination also guarantees a relatively low noise impact. Unfortunately, scanners are not PSF-free devices and the value or color of any pixel is bled to the surrounding pixels. This well-known constraint limits the practical resolution of microscopes, telescopes and digital cameras. What role does the PSF play in the recovery of HH? The effect is blurring the ColHis. There is no easy way to fight against PSF, but its effects can be minimized following a strategy in which similar pixels are placed together. This takes us back to the origination stage.

[ii] $\Delta E = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2}$

Perfect ordering of a relatively small HH, containing 400 tiles, can be an impossible task using a brute-force strategy because the number of possible arrangements (≈400!) is a 900 figures number. There is no analytic or reasoned algorithm to search for the optimum solution. This specific problem is part of the so called "global optimization" branch in Mathematics. Out of the many available possibilities, a meta-heuristic approach known as *simulated annealing*[6] has been used for both simplicity and potential. In brief, the iterative algorithm takes an arbitrary starting tile distribution and exchanges two elements looking for a better solution; in case the exchange worsens the solution it is sometimes accepted in the quest of an overall better solution. It is originally inspired in metallurgy, in the annealing strategies followed with real alloys to achieve higher quality crystals.

The Figure 9 summarizes what would be the basic cycle of a HH.

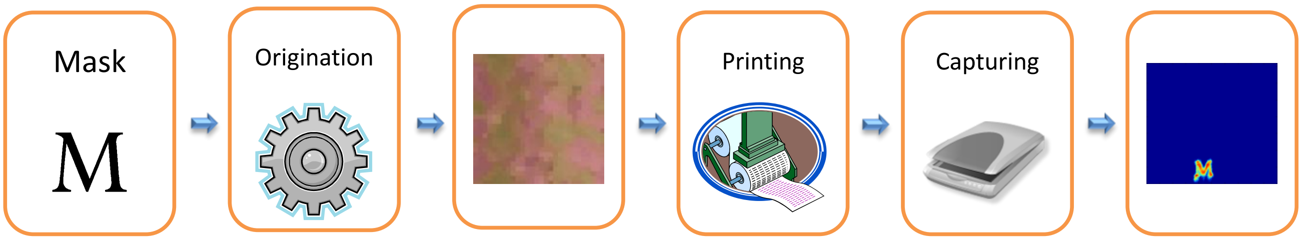



Figure 9. Basic life-cycle of a hidden histogram

## 6. FURTHER PROGRESS IN HIDDEN HISTOGRAMS: HISTOCODES AND DYNAMIC CALIBRATION

Two lines of progress will be sketched in this section. We will not delve into technical details in order to keep the scope of the paper reasonable, furthermore it is an area under active development. The first one is related to dynamic calibration of capture devices, opening the door to on-the-fly calibration, ideally transparent to the end user. The second one is related to embedding of digital information instead of human-readable shapes.

In the previous section, the capture device used has been a scanner, properly calibrated with an ICC profile. Obtaining an ICC profile in a scanner usually involves capturing a color chart whose colors have been previously measured with an spectrophotometer[7]. Same approach could be used with a digital camera, even a mobile phone camera. The main difference in the latter case is that calibration is strongly linked to the illuminant used. Different illuminants or changes in the color temperature of the illuminant make it necessary a new calibration and digital cameras usually take pictures in different light conditions. For a non-trained end-user, the ideal scenario would be that in which calibration is performed on the fly, with information embedded in the ColHis. The idea is depicted in Figure 10, where a number of tiles in the HH are reserved for calibration purposes. A pseudo-random distribution will give the maximum protection against local defects. The software analyzing the captured image would need to know in advance the location and value of each calibration tile. Note that this forces to use a custom calibration procedure for the camera/scanner.

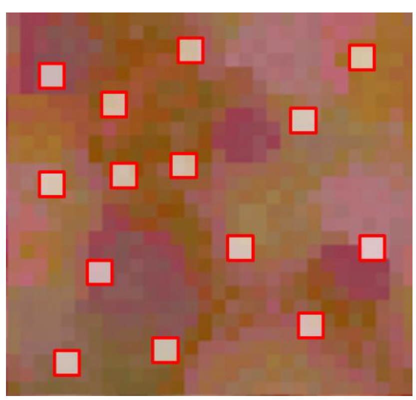


Figure 10. Example showing that a percentage of the tiles in the image are reserved for calibration purposes. The pixels in the surroundings accommodate to minimize the color differences.

One of the biggest strengths of HH, together with the simplicity of the hardware required, is the counterfeit resilience. Internal tests have been carried out at early stages, showing in general extraordinary robustness to simple cloning of a genuine document (needless to say re-origination). Preliminary independent tests have confirmed this fact. If we can fabricate unique printed documents whose information is very difficult to clone, the security feature can be used not only with authentication purposes, but also with integrity purposes: the personalization data of an ID document, for example. One way to achieve this is by using 2D barcodes in the masks instead of human-readable shapes. The concept of embedding bits in the HHs is what we call Histocodes. Commercial solutions hindering document cloning, such as Ingenia laser surface authentication[8] or Prooftag Bubble Tag[9], aim the same objectives with different clever solutions.

The Figure 11 shows an example of a Histocode carrying a 25 bits code in the *Lh* and another 25 bits code (which could be identical) in the *ch*. The original image has 3.3 cm width and the reproduction was produced using a laser copier. The calibration information was embedded in the image, so any non-calibrated device could have been used. A scanner was used but a digital camera achieves comparable results. The *Lh* and *ch* histograms have been zoomed in to better appreciate the alignment of the bits with the grid. What can be observed is that the reproduction process blurs the histogram information and makes practically impossible to recover the digital codes.

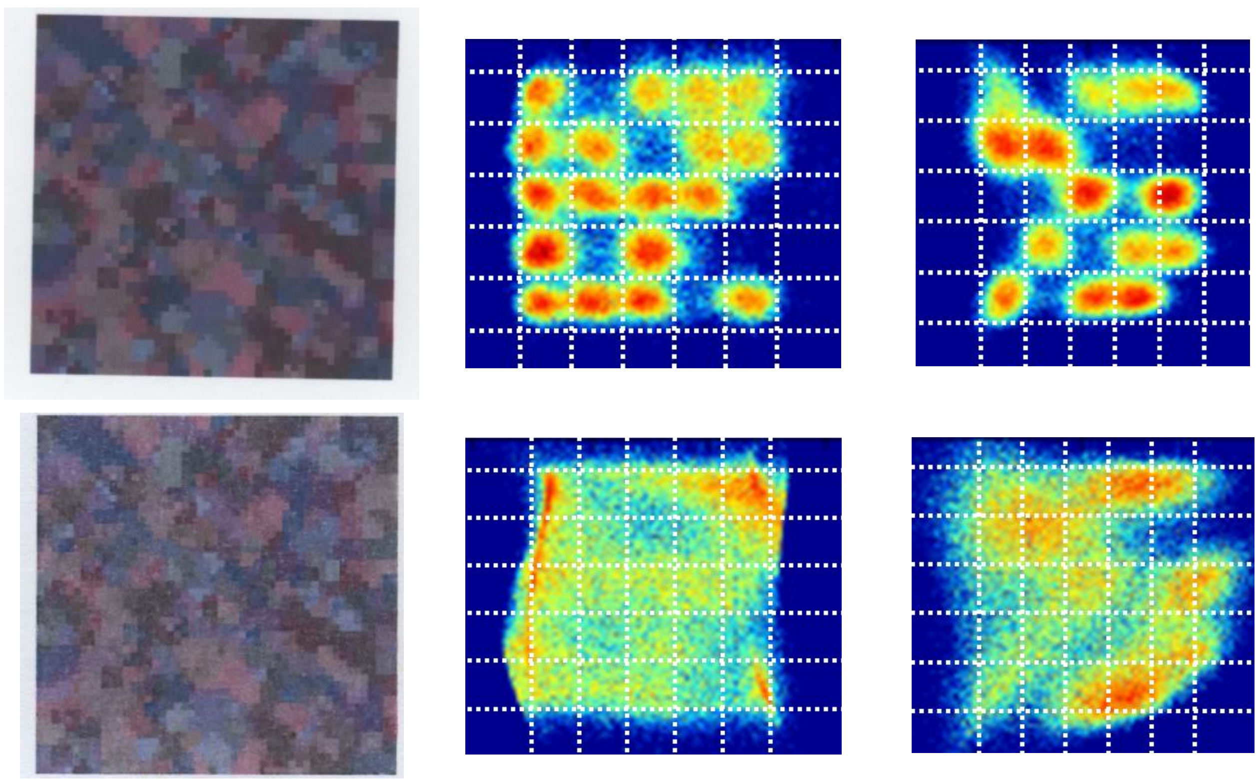


Figure 11. Histocode image (left column), with calibration points embedded in the image, and two different (simultaneous) 25 bits codes: middle column the *Lh* histogram and right column the *ch* histogram. The upper row shows the information for the genuine document and the lower one for the reproduction.

As a final thought, the dynamic calibration embedded in this Histocode has an almost infinite number of combinations (position and order of the calibration points) for the same 25+25 bits code, so additional digital information could be used to further complicate the recreation of a Histocode.

## 7. CONCLUSIONS

Bivariate color histograms are a powerful procedure to create color fingerprints of documents. Due to its intrinsic statistical nature they provide accuracy on the average values but tolerance on individual errors at printing and capture.

Color histograms are a valuable tool to provide, probably first time ever, color control on the whole printed document (instead of lateral control patches). This has enormous potential on quality control, counterfeit detection or soiling detection.

Hidden histograms are printed security features in which the color histogram of the document reveals up to two simultaneous hidden patterns.

The strength of Hidden histograms relies on a very complex origination process and advanced calibration techniques for printing and digital capture. Complexity is restricted to the software not to the hardware, professional off-the-shelf inkjet printers and desktop capture devices were used.

Direct reproduction implies an additional capture + reprint color process whose deformations can be easily identified. This makes it especially suitable to encode digital information which could guarantee integrity and authenticity of a document/information.

## 8. REFERENCES


1. A. Sharma, *Understanding Color Management*, Thomson Delmar Learning, p. 80, 2004.
2. Colorthink Pro by Chromix Inc., http://www.chromix.com/colorthink/
3. Matlab by Mathworks, http://www.mathworks.com
4. Bivariate analysis, Wikipedia, http://en.wikipedia.org/wiki/Bivariate_analysis
5. J. Angulo, "Modelling and segmentation of colour images in polar representations", *Image and Vision Computing 25*, pp. 475-495, 2007.
6. S. Kirkpatrick, C.D. Gelatt and M.P. Vecchi, "Optimization by Simulated Annealing", Science, vol. 220, n. 4598, pp. 671-680, 1983.
7. B. Fracer, C. Murphy and F. Bunting, *Real World Color Management* (2nd edition), Peachpit Press, Berkeley, Chapter 7, 2005.
8. LSA™ by Ingenia, http://www.ingeniatechnology.com/
9. Bubble Tag™ by Prooftag™, http://www.prooftag.net