Experimental study of Print-and-Scan impact as random process

Iuliia Tkachenko, William Puech, Olivier Strauss, Christophe Destruel

DOI: 10.2352/ISSN.2470-1173.2016.8.MWSF-089

Experimental study of Print-and-Scan impact as random process

Iuliia Tkachenko [1,2]***, William Puech*** [1]***, Olivier Strauss*** [1]***, Christophe Destruel*** [2]
[1] ***LIRMM, UMR CNRS 5506, University of Montpellier, France***
[2] ***Authentication Industries, Montpellier, France***

Abstract

The Print-and-Scan (P&S) process is generally considered as being a random process that can be modeled by a white additive normal process, ergodic in the wide sense. This study aims at experimentally validate or invalidate this hypothesis. Moreover, the experiments we carried on have been conducted in order to separate the printing from scanning impact. The main conclusion of these experiments is that the usual hypothesis cannot be supported by the experiments. The normal modeling can be invalidated for the whole P&S process. Moreover, it has been highlighted that this process cannot be considered as being neither white nor ergodic in the wide sense. The scanner noise seems to be mean ergodic and closer to a Laplace, than to a normal distribution.

Introduction

The degradation of information due to Print-and-Scan (P&S) processes is a major issue in digital forensics and printed document as well as in image authentication. This degradation is usually considered as being a stochastic process that can be modeled by an additive or a multiplicative normal or lognormal distribution [1, 2]. These models include a random noise due to ink dispersion on the paper during the printing as well as the illumination conditions during the scanning process. It is generally acknowledged that these two degradations cannot be separated.

In this paper, we aim at experimentally determine the nature of the printer and the scanner noise, as well as identify the distribution of white and black pixels after the P&S process.

One of the main goals of this paper is to propose a way to characterize the nature of a stochastic process in image processing and particularly answer the following questions:

1. Does this process follow a given statistical distribution (e.g. normal or Laplace)?
2. Can we consider the noise as being additive?
3. Can we consider the noise as being stationary?
4. Can we consider the noise as being ergodic?
5. Can we consider the noise as being white?

To answer those questions, we propose a series of statistical tests. We experiment those tests on a large image database that contains black-and-white images of textured image [4] that have been collected after numerous printing and scanning operations. We use these images since they have a very high contrasted structure made of black and white patterns. All statistical tests presented in this paper are performed with real data.

In the next section, we introduce some statistical definitions and explain how the statistical tests we use work. Then we propose different methodologies to verify weather the considered process is stationary and ergodicy or not. An overview of P&S models, as well as the description of textured image we use are discussed. The experiments and their outcomes are presented in following section. The experimental results we obtain on color changes after P&S process are discussed. Finally, we conclude and determine several future paths in the last section.

Statistical definitions and tests

In this section we present the characteristics of random process as well as several hypothesis test, that are used for our experiments.

Random process characteristics

A random variable, $X(s)$, is a single-valued real function that assigns a real number, called the value of $X(s)$, to each sample point $s \in S$ [3].

A random process can be defined as a family of random variables $\{X(t,s)|t \in T, s \in S\}$ defined over a given probability space and indexed by the time parameter t [3]. The $X(t,s)$ is a collection of time functions, one for each sample point s (see Fig. 1).

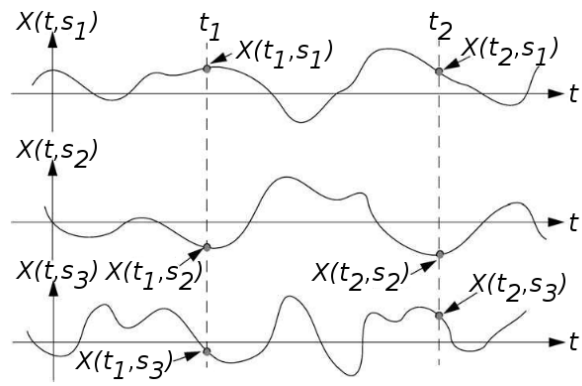


Figure 1: A sample random process [3].

A parameter is called statistical, if it is calculated using $X(t,s)$ with a fixed value of t. A parameter is called spatial, if it is calculated using $X(t,s)$ with a fixed value of s.

Stationary process. A random process is called a *strict sense stationary process* if its Cumulative Distribution Function (CDF) F_X is invariant to a shift in the time origin [3]. That means that $X(t,s)$ is strict sense stationary if its CDF is identical to the CDF of $X(t+\varepsilon,s)$ for any arbitrary shift ε:

$$F_X(X(t,s)) = F_X(X(t+\varepsilon,s)). \quad (1)$$

When the CDF is differentiable, the equivalent condition for strict sense stationarity is that the Probability Distribution Function (PDF) f_X is invariant to a shift ε in the time origin:

$$f_X(X(t,s)) = f_X(X(t+\varepsilon,s)). \quad (2)$$

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.8.MWSF-089

In practice, we often work only with the mean and the autocovariance functions of a random process. A random process in which the mean and autocovariance function does not depend on absolute time is called a *wide-sense stationary* process. Thus, for a wide-sense stationary process $X(t)$, we have:

$$E[X(t,s)] = \mu_X, \quad \text{constant}, \tag{3}$$

$$E(X(t,s),X(t+\tau,s)) = R_{XX}(\tau), \forall t \in T. \tag{4}$$

Ergodic process. The time average of random process $X(t,s)$ is calculated at a fixed sample point s (i.e. calculated over $X(t_1,s),\ldots,X(t_n,s)$). Considering a random process $X(t,s)$ whose observed sample is $x(t)$, the time average of the function $x(t)$ is defined by:

$$\bar{x} = \lim_{\tau\to\infty} \frac{1}{2\tau} \int_{-\tau}^{\tau} x(t)dt. \tag{5}$$

A stationary random process $X(t,s)$ is said to be *ergodic* if every member of the set exhibits the same statistical behavior as the set. This implies that it is possible to determine the statistical behavior of the set by examining only one typical sample function [3]. An ergodic process can be represented by only one stochastic process realization. As for stationarity, ergodicity is often restricted to its two first orders.
A random process $X(t,s)$ is defined as being *first order ergodic* (or ergodic in the mean) if time mean function has the same distribution as statistical mean function:

$$\bar{x} = \mu_X. \tag{6}$$

A random process $X(t,s)$ is defined to be *second order ergodic* if time autocovariance function has the same distribution as statistical covariance function:

$$R_{XX} = Cov_{XX}. \tag{7}$$

White-noise process. The random process $X(t,s)$ is a white noise if its autocovariance function is close to a Dirac impulse:

$$R_{XX}(\tau) = \frac{\sigma_X^2}{2}\delta(\tau), \tag{8}$$

where σ_X^2 is the variance of the random process $X(t,s)$ and $\delta(\tau)$ is the impulse function.
The Fourier transform (if it exists) of the autocovariance function of a stationary in the wide sense random process $X(t,s)$ is called its power spectral density. Since $R_{XX}(\tau) = \frac{\sigma_X^2}{2}\delta(\tau)$, then $S_{XX}(\tau) = TF\{R_{XX}(\tau)\} = \frac{\sigma_X^2}{2}$ does not depend on the frequency. This is why it is called white noise (by analogy with the spectral property of the white light).

χ^2 goodness of fit test

The χ^2 goodness of fit test is used to identify whether sample data are consistent with a given distribution or not [13]. This test can be used when the sample data are categorical and when the number of observations in each variable level is at least 5.
The null hypothesis of this test is: *the data are consistent with the specified distribution*. The alternative hypothesis is: *the data are not consistent with the specified distribution*. Generally, only one reject of null hypothesis is enough to ensure, at a given significant level, that the data are not consistent with the distribution. However, several tests are needed to be confident that the null hypothesis can be accepted at the same significance level. The significance level is chosen by the user. It is often equal to $0.01, 0.05$, or 0.10.
Let X be a discrete random variable, whose domain can be divided into k partition classes $A_1, A_2, \ldots, A_k$. Let $n_i, i = 1,\ldots,k$ be the number of samples in each class i, with $N = n_1 + n_2 + \cdots + n_k$ being the total number of samples in A. Let $p_i, i = 1,\ldots,k$ be the probabilities of the class i based on the specified distribution. The χ^2 goodness of fit test considers the statistics D^2 defined as follow:

$$D^2 = \sum_{i=1}^{k} \frac{(n_i - Np_i)^2}{Np_i}. \tag{9}$$

The degree of freedom of χ^2 goodness of fit test equals to $k-1$. The null hypothesis can be accepted if:

$$D^2 \sim \chi^2. \tag{10}$$

The table of critical values for the χ^2 goodness of fit test presents threshold values $\chi^2_{k-1,\alpha}$ for this test for different significance levels α (also called p-value). The null hypothesis is rejected if:

$$D^2 > \chi^2_{k-1,\alpha}. \tag{11}$$

If the same A is used to estimate l parameters of the specified distribution, then the $\chi^2_{k-1-l,\alpha}$ critical value has to be chosen.

Kolmogorov-Smirnov test

The Kolmogorov-Smirnov (KS) test can be used to verify whether two probability distributions differ or not [13]. The null hypothesis of this test is: *the two distributions are close*. The alternative hypothesis is: *the two distributions are different*. Acceptance or rejection are also done at significance level chosen by user.
This test uses the maximal distance between the empirical distribution functions of both probability distributions. The empirical distribution function F_n for n observations x_i of X is defined as:

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{]-\infty,x]}(x_i), \tag{12}$$

where $I_{]-\infty,x]}(x_i)$, the indicator function of $]-\infty,x]$, equals to 1 if $x_i \leq x$ and to 0 otherwise.
Let X_1 and X_2 be two discrete random variables. Let $F_{1,n}$ and $F_{2,n'}$ be the empirical distribution functions of random variables X_1 and X_2 respectively. The KS test considers the statistics $D_{n,n'}$ defined as follow:

$$D_{n,n'} = \sup_x |F_{1,n}(x) - F_{2,n'}(x)|. \tag{13}$$

The null hypothesis is rejected at significant level α if:

$$D_{n,n'} > c(\alpha)\sqrt{\frac{n+n'}{nn'}}, \tag{14}$$

where $c(\alpha)$ is defined using the table of critical values for the Kolmogorov-Smirnov test [14]. For example, $c(\alpha) = 1.36$, if $\alpha = 0.05$ or $c(\alpha) = 1.22$, if $\alpha = 0.1$.

DOI: 10.2352/ISSN.2470-1173.2016.8.MWSF-089

Mann-Whitney test

The Mann-Whitney U-test (also called Wilcoxon rank-sum test) is a non-parametric test that is used to test whether two independent random variables have identical distributions or not [13]. The null hypothesis of this test is: *the two distributions are identical*. The alternative hypothesis is: *the two distributions are different*. As in the previous hypothesis tests, acceptance or rejection are done at a user chosen significance level.

This test is based on the idea that the information about the relationship between two random variables X_1 and X_2 can be obtained from n_1 observations of the random variable X_1 and n_2 observations of the random variable X_2 that are arranged together in increasing order.

The test aims at testing whether the two random variables X_1 and X_2 are mixed or not. Let us suppose that the random variable X_1 has less observations, e.g. $n_1 < n_2$. The combined set of data is first arranged in ascending order with tied scores receiving a rank equal to the average position of those scores in the ordered sequence. Then, the ranks for the observations which came from X_1 are summed up to provide the rank sums R_1. The Mann-Whitney U-test considers the statistics U defined as follow:

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1. \tag{15}$$

The test consists of comparing the value U with the value given in the table of critical values for the Mann-Whitney U-test $U_{n_1,n_2,\alpha}$, where the critical values are provided for given values n_1, n_2 and α. The null hypothesis is rejected if:

$$U > U_{n_1,n_2,\alpha}. \tag{16}$$

Proposed methodologies

In this section, we present how to use the proposed tests to verify the random process stationarity and ergodicity of first order and its stationarity of second order. Let us consider the discrete random process $X(t,s)$ defined in finite sequences $t = (t_1, \cdots, t_m)$ and $s = (s_1, \cdots, s_n)$.

Stationarity of first order test

As it was mentioned before, the process is stationary of first order if its statistical average $E[X(t,s)]$ is constant over the temporal separation t, i.e. $E[X(t,s)] = \mu_X$.

To verify the first order stationarity of random process $X(t,s)$, we calculate its statistical mean values for each time t_i, i.e. we have a set of m mean values. In order to show the constancy of statistical mean, we divide this set into two independent random subsets and apply the Kolmogorov-Smirnov test. The null hypothesis of this test is formulated as: *the two independent random subsets of mean value set have the same distributions*.

If the null hypothesis is rejected at significant level α at least once, the random process $X(t,s)$ is not stationary of first order. Otherwise, the random process $X(t,s)$ can be considered as being stationary of first order.

Ergodicity of first order test

The process is ergodic of first order if the spatial mean equals the statistical mean. As the random process is stationary of first order, we have the set of m statistical mean values. Then we calculate the spatial mean value for each state s_j (thus we obtain a set of n spatial mean values).

To verify the first order ergodicity of random process $X(t,s)$, we perform a Mann-Whitney U-test. The null hypothesis is formulated as: *the spatial mean and statistical mean are identical*.

If the null hypothesis is rejected at significant level α, the random process $X(t,s)$ is not ergodic of first order. Otherwise, we declare that the random process $X(t,s)$ can be considered as being ergodic of first order.

Stationarity of second order test

We want to verify whether the random process $X(t,s)$ is stationary of second order or not. As it was mentioned before, the process is stationary of second order if the autocovariance function does not depend on absolute time, i.e.

$$E(X(t,s), X(t+\tau,s)) = R_{XX}(\tau), \forall t \in T.$$

To verify the second order stationarity, we calculate its autocovariance function at each time t (i.e. we have a set of m' autocovariance values). In order to show its constancy, we divide this set of autocovariance values into two independent random subsets and apply the Kolmogorov-Smirnov test. The null hypothesis of this test is formulated as: *the two independent random subsets of autocovariance set have the same distributions*.

If the null hypothesis is rejected at significant level α at least once, the random process $X(t,s)$ is not stationary of second order, and therefore, it is neither wide-sense stationary nor wide-sense ergodic. Otherwise, we declare that the random process $X(t,s)$ can be considered as being stationary of second order.

P&S impact

In this section we present the state of the art of P&S process modeling as well as the description of images used for our experiments.

P&S process modeling

Here we review some previous interesting work dedicated to characterize the P&S process.

The purpose of a printer model is to accurately predict the gray level of a binary image produced by a printer. The authors in [5] suggest such a printer model that can be used by halftoning algorithms. They use a previously proposed physical model to train the adaptive signal processing model offline. Then this model is used to calculate the average exposure of each subpixel for any input pattern in real time.

In [6], a print-quality perception model has been proposed. This model uses an image analysis system and a neural network trained to be able to differentiate different print qualities.

The authors in [7] model the P&S process by considering the pixel value and the geometric distortions separately. According to this model, the distortion of pixel values is caused by the luminance, contrast, gamma correction, chrominance variations and blurring of adjacent pixels. This distortion introduces a visual quality change into the scanned image. The distortion of the geometric boundary is caused by rotation, scaling and cropping. That may introduce considerable changes at the signal level, especially on the DFT coefficients.

The pixel value distortion model for inkjet printers and flatbed scanners, proposed in [7], consists of a high-pass filter like a point

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.8.MWSF-089

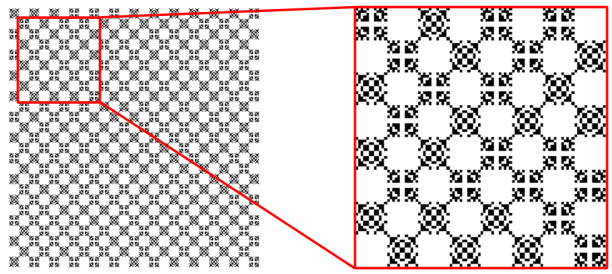
spread function, a white normal random noise, a thermal noise and a dark current noise. The size of P&S image is usually different from the original, as printer and scanner resolutions may be different [7]. The authors consider the image geometric distortions to be an important factor and view it as an additional source of noise.
In [8], the authors view the image cropping as a cause of blurring in frequency domain. They model the P&S process on three main components: mild cropping, correlated high-frequency noise and nonlinear effects. The conclusions of the authors, based on observing the DFT coefficient magnitudes during their experiments for laser printers, are as follows:

- The low and mid frequency coefficients are less sensitive to P&S process than the high frequencies,
- The coefficients of the low and mid frequency bands with low magnitudes suffer from a much higher noise than their neighbors with high magnitudes,
- Coefficients with higher magnitudes have a gain of roughly unity,
- Slight modifications of the high magnitude low frequency coefficients do not add significant distortion to the image.

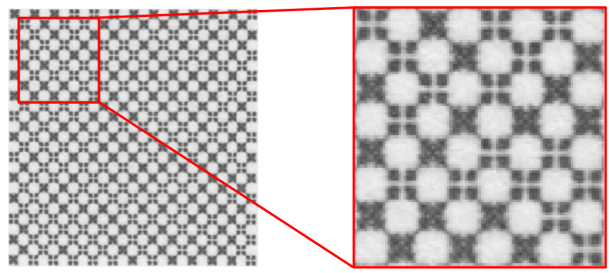
These observations suggest that the printing operation does not cause blurring, since several dots are used to print each pixel of a digital image [8].
As the halftone introduces distortions in the P&S image too, the authors in [9] proposed an accurate linear model for error diffusion during halftoning. This model predicts the high-frequency noise and the edge sharpening effects introduced by halftone. The authors in [10] use this halftone model to represent halftone/inverse halftone printing channel by an equivalent channel. This model was constructed for inkjet printer by supposing that the scanning process does not introduce any distortion.
Some researchers model the P&S channel as an authentication channel [11]. The printing process at a very high resolution can be seen as a stochastic process due to the nature of the printer characteristics [12]. The authors simulate the printing process as a generalized normal distribution or a log-normal distribution (proposed by [1]). Assuming the statistical properties of the P&S channel, the authors in [1] experimentally show that the metric values of black/white blocks of pixels are related to random variables following an asymmetric log-normal distribution.
The modeling of P&S process is a well-known challenging field. Due to difficulties and randomness of this process, according to our expertise, no suitable mathematical model has been suggested yet. Meanwhile, some interesting earlier work are presented in this section.

Textured image description

In order to visualize the impact of the P&S process, we printed and scanned the same textured image several times. The original textured image is illustrated in Fig. 2.a and its printed and scanned in 600 dpi version is illustrated in Fig. 2.b. The textured patterns have a size of 12×12 pixels. Their characteristics are described in [4]. From the zoomed parts of these samples (Fig. 2), we note that the internal textured structure of the patterns is lost, the image becomes blurred and the colors are changed.
As mentioned before, these textured images will be used in our experiments due to their frequent color changes and also to the fact that the texture influences the color distortion.



(a) Textured image



(b) Printed and scanned textured image

Figure 2: Examples of a) Original textured image and b) Printed and scanned textured image (a).

It is possible to compensate some color changes by using a color Look-Up-Table (LUT), dedicated to each printer-scanner pair. Such a LUT can be constructed by measuring the color changes after a P&S process. An example of such a LUT is presented in Fig. 3.

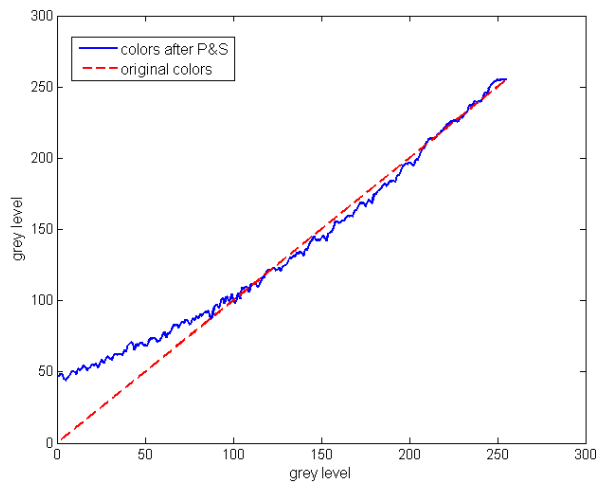


Figure 3: Example of color changes after P&S process. Red line: original colors, blue line: colors after P&S process.

We note from Fig. 3 an especial loss in dark colors. The LUT can correct some color defects. However, it is impractical to construct such a LUT for each printer-scanner pair and each type of paper.

DOI: 10.2352/ISSN.2470-1173.2016.8.MWSF-089

Experiments: noise after P&S impact

In this section, we present a series of experiments dedicated to the study of the nature of the noise added by a P&S process. First, we isolate the scanner noise and perform several statistical tests to study its nature and to investigate its characteristics. Then, we perform the same experiments to study the noise added by the whole P&S process.

Study of scanner noise

Experimental setup. This experiment aims at isolating the scanner noise. To achieve this isolation, we propose to print an image once and then scan it n times. Let I be the original image, P be the printed image of original image I and S_j $(j = 1,\dots,n)$ be the n scanned samples of printed image P. The scheme can be picked as:

$$I \rightarrow \textbf{Print} \rightarrow P \rightarrow \textbf{Scan} \rightarrow S_j. \tag{17}$$

Let us consider P as a function of I and ε_P, S_j as a function of I, ε_P and ε_S, i.e:

$$P = f(I, \varepsilon_P),$$

$$S_j = g(P, \varepsilon_{S_j}) = P + \varepsilon_{S_j} = f(I, \varepsilon_P) + \varepsilon_{S_j},$$

where $f()$ is the function associated to the printing process, $g()$ is the function associated to the scanning process and ε_P and ε_{S_j} are the noises introduced by the printer and the scanner respectively. The $+$ operator shows that the printed and scanned image S_j can be represented by a function of the printing process and the scanner noise (that also depends on the function of the printing process).
If the noise has a regularity, we can consider to calculate the difference among samples in order to study the noise nature. We propose thus to calculate the differences among each pair of n samples, i.e. subtract pairs of scanned images to provide samples of the scanner noise:

$$S_j - S_{j'} = \varepsilon_{S_j} - \varepsilon_{S_{j'}}, \tag{18}$$

The $\varepsilon_{jj'} = \varepsilon_{S_j} - \varepsilon_{S_{j'}}$ can be used to characterize the stationarity and the ergodicity of the noise introduced by the scanning process. These characteristics, that are inherent to $\varepsilon_{jj'}$, are also inherent to ε_{S_j} and $\varepsilon_{S_{j'}}$. Therefore, we can characterize the scanner noise by characterizing the stationarity and ergodicity of $\varepsilon_{jj'}$.

Let us introduce the same experimental setup when using the terminology of random processes. After a P&S process each image can be considered as random variable $X(t,s)$. Therefore, the set of n P&S images can be considered as random process $X(t,s)$, where $s = \{1,\dots,n\}$ is the dimension of our data set and $t = \{1,\dots,m\} \times \{1,\dots,m\}$ is the set of pixels in an image I.
As shown before, the scanner noise function can be obtained by subtracting every two random variables $X(t,s_j)$ and $X(t,s_{j'})$, in order to create a new random process $Y(t,s)$ with $s = \{1,\dots,n'\}$:

$$Y(t,s) = X(t,s_j) - X(t,s_{j'}), j \neq j', s_j = s_{j'} = \{1,\dots,n'\}, \tag{19}$$

where $n' = \frac{n(n-1)}{2}$.

Database description. We print one textured image once and then scan it $n = 90$ times. Each P&S image has a 300×300 pixel size, i.e. $m = 300$. The number of noise samples is then equal to $360,450,000$.
Let us visualize the distribution of $Y(t,s)$ for any values of t and s (i.e. as if it was an ergodic process) in Fig. 4. The mean value of this distribution is equal to $\mu = -0.0187$, its standard deviation is equal to $\sigma = 10.2264$.

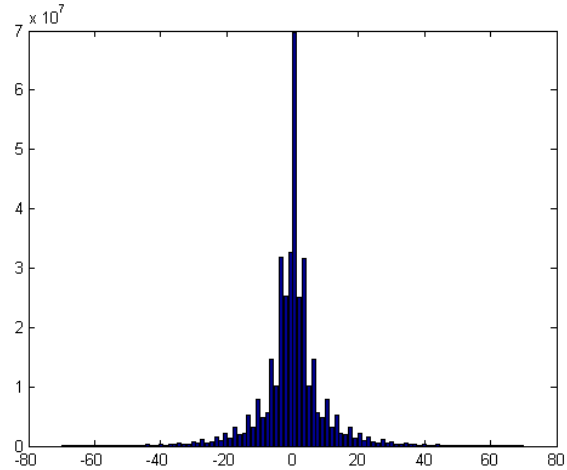


Figure 4: Scan noise distribution.

Characterization of the $Y(t,s)$ distribution. In this experiment, we want to verify whether the $Y(t,s)$ follow a given classical distribution (a normal or a Laplace) or not. In order to answer this question, we compare the distribution of the random process $Y(t,s)$ with each specified distribution using the χ^2 goodness-of-fit test described previously.
A reduced sample data set has been created by randomly selecting $n' = 100$ vectors of $Y(t,s)$ (for lower computational time). We therefore work with a data subset of $9,000,000$ samples. To counteract working with a subsample, we perform each test several times.
An algorithm of χ^2 goodness-of-fit test involving the estimation of mean and variance is presented in Algorithm 1. The result of this test is that both hypotheses are rejected at a significance level of 0.05. Thus our data-set has neither a normal nor a Laplace distribution.
We computed two distances to see how far the empirical CDF based on the initial data is from the CDF of a normal or of a Laplace distribution. The comparison involves:

- the maximal distance d_{max} between the CDFs (the measure used in Kolmogorov-Smirnov test):

$$d_{max} = \sup_x |F_n(x) - F(x)|, \tag{20}$$

- the squared distance d^2 between the CDFs (the measure used in Cramer-von Mises test):

$$d^2 = \sum_i (F_n(x_i) - F(x_i))^2 (x_i - x_{i-1}). \tag{21}$$

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.8.MWSF-089

Algorithm 1 Chi-square goodness-of-fit test

Require: Data $X = \{x_1, \ldots, x_N\}$
1: Calculate the data histogram h with k bins
2: n_i number of samples in bin i
3: $\mu \leftarrow mean(X)$
4: $\sigma \leftarrow std(X)$
5: **for** $i = 1 : k$ **do**
6: $p_i \leftarrow \int f(x, \mu, \sigma)dx$
7: **end for**
8: $D^2 \leftarrow \sum_{i=1}^{k} (\frac{(n_i - Np_i)^2}{Np_i})$
9: **if** $D^2 < \chi^2_{k-3}$ **then**
10: X has distribution, that corresponds to probability density function f
11: **end if**

Table 1 presents the differences among the CDF based on our data-set and the CDF(s) of a normal and a Laplace distributions. We notice that these distances are rather stable under the changes of sample size.

Number of samples N	Laplace distibution		Normal distibution	
	d_{max}	d^2	d_{max}	d^2
30,000	0.0933	0.0712	0.1482	0.2507
100,000	0.0936	0.0722	0.1483	0.2496
200,000	0.0922	0.0704	0.1474	0.2482
300,000	0.0934	0.0719	0.1490	0.2488
400,000	0.0925	0.0714	0.1484	0.2497

Table 1: Distances among CDFs of initial data (scanner noise) and estimated Laplace and normal distributions.

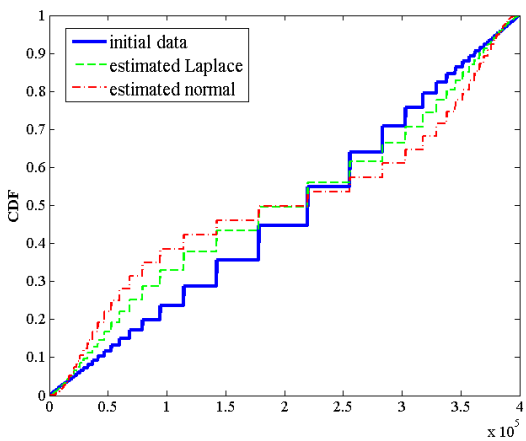
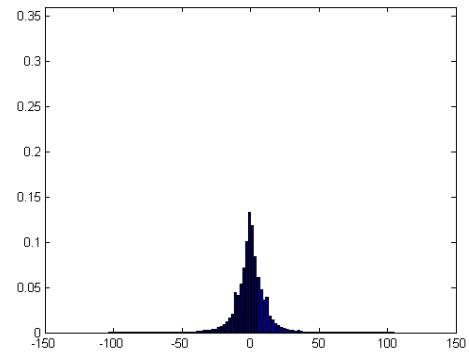


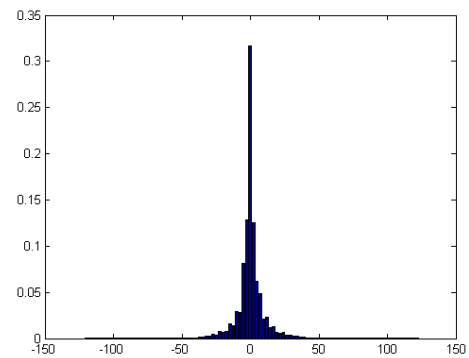
Figure 5: CDF of scanner noise (blue line), estimated Laplace (green line) and estimated normal (red line) distributions.

We illustrate these distances among CDFs in Fig. 5. We note that the CDF of the initial data (blue line) is closer to the CDF of the Laplace distribution (green line) than the CDF of the normal distribution (red line). Therefore the scanner noise can be said to be closer to a Laplace distribution than to a normal distribution.

Additive noise. After a P&S process, the image obtained with one of the textured images consists of gray level pixels when the original textured image is binary. Therefore, all gray level pixels can be separated into two classes: the black class, that consists of black pixels from textured image, and the white class, that consists of white pixels from textured image.
The noise cannot be assumed to be additive if the error distribution of the black pixels is different from the error distribution of the white pixels. As we know the true map of black and white pixel placement, we separate the $\varepsilon_{jj'}$ in two classes: black class and white class. This leads to two random processes $B(t,s)$ and $W(t,s)$. The histograms of error black random processes and white random processes are illustrated in Fig. 6.
We apply the Mann-Whitney test in order to decide whether or not the error distribution can be said to be additive, i.e. quantify the hypothesis that the black and white distributions are the same. The hypothesis that the distributions are the same is rejected at a significance level of 0.05. Therefore, the noise added by the scanning process cannot be considered as being additive.



(a) Black pixel class



(b) White pixel class

Figure 6: Histograms of scan noise introduced to a) Black pixels and b) White pixels.

Stationarity of first order. To verify the first order stationarity of random process $Y(t,s)$, we use the method described previously. The null hypothesis of Kolmogorov-Smirnov test is formulated as: *the two independent random subsets of mean value set have the same distributions*. The Kolmogorov-Smirnov test cannot reject the null hypothesis at a significance level of 0.05, therefore our process can be supposed being stationary of first order, i.e. the statistical average respects $E[Y(t,s)] = \mu_Y$.

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.8.MWSF-089

Ergodicity of first order. To verify the first order ergodicity of the random process $Y(t,s)$, we use the method described previously. The null hypothesis is formulated as: *the spatial mean and statistical mean are identical*. The Mann-Whitney test cannot reject the null hypothesis at a significance level of 0.05. Therefore, our process can be supposed being ergodic of first order.

Stationarity of second order. To verify the second order stationarity of the random process $Y(t,s)$, we use the method described previously. The null hypothesis of the Kolmogorov-Smirnov test is formulated as: *the two independent random subsets of autocovariance set have the same distributions*. The Kolmogorov-Smirnov test rejects the null hypothesis at a significance level of 0.05, therefore our process is not stationary of second order. That means our process is not wide-sense stationary.

Test for white noise. A random process can be a white-noise process if it is wide-sense stationary. Our process is not wide-sense stationary, therefore our process is not a white-noise process.

Study of P&S noise

Experimental setup. This experiment aims at studying the noise added by P&S process. To achieve this noise, we propose to print an image n times and then scan each printed image once. Let I be the original image, P_i $(i = 1,\dots,n)$ be the n printed images of original image I and S_i be the n scanned samples of printed images P_i. The scheme can be picked as:

$$I \rightarrow \textbf{Print} \rightarrow P_i \rightarrow \textbf{Scan} \rightarrow S_i. \tag{22}$$

Let us consider P_i as a function of I and ε_{P_i}, S_i as a function of I, ε_{P_i} and ε_{S_i}, i.e:

$$P_i = f(I, \varepsilon_{P_i}),$$
$$S_i = g(P_i, \varepsilon_{S_i}) = P_i + \varepsilon_{S_i} = f(I, \varepsilon_{P_i}) + \varepsilon_{S_i},$$

where $f()$ is the function associated to the printing process, $g()$ is the function associated to the scanning process and ε_{P_i} and ε_{S_i} are the noises introduced by the printer and the printer-and-scanner respectively. The $+$ operator shows that the printed and scanned image S_i can be represented by a function of the printing process and the scanner noise (that also depends on the function of the printing process).
If the noise has a regularity, we can consider to calculate the difference among samples in order to study the noise nature. We propose thus to calculate the differences among each pair of n samples, i.e. subtract pairs of scanned images to provide samples of the P&S noise:

$$\begin{aligned} S_i - S_{i'} &= \varepsilon_{S_i} - \varepsilon_{S_{i'}} \\ &= \varepsilon_P + \varepsilon_{PS_i} - \varepsilon_P - \varepsilon_{PS_{i'}} \\ &= \varepsilon_{PS_i} - \varepsilon_{PS_{i'}}, \end{aligned} \tag{23}$$

The $\varepsilon_{ii'} = \varepsilon_{PS_i} - \varepsilon_{PS_{i'}}$ can be used to characterize the stationarity and the ergodicity of the noise introduced by the P&S process. These characteristics, that are inherent to $\varepsilon_{ii'}$, are also inherent to ε_{PS_i} and $\varepsilon_{PS_{i'}}$. Therefore, we can characterize the P&S noise by characterizing the stationarity and ergodicity of $\varepsilon_{ii'}$.

Let us introduce the same experimental setup when using the terminology of random processes. After a P&S process each image can be considered as random variable $X(t)$. Therefore, the set of n P&S images can be considered as random process $X(t,s)$, where $s = \{1,\dots,n\}$ is the dimension of our data set and $t = \{1,\dots,m\} \times \{1,\dots,m\}$ is the set of pixels in an image I.
As shown before, the P&S noise function can be obtained by subtracting every two random variables $X(t,s_j)$ and $X(t,s_{j'})$, in order to create a new random process $Y(t,s)$ with $s = \{1,\dots,n'\}$:

$$Y(t,s) = X(t,s_j) - X(t,s_{j'}), j \neq j', s_j = s_{j'} = \{1,\dots,n'\}, \tag{24}$$

where $n' = \frac{n(n-1)}{2}$.

Database description. We print one textured image $n = 30$ times and then scan all printed images once. Each P&S image has a 300×300 pixel size, i.e. $m = 300$. The number of noise samples is then equal to $39,150,000$.
Let us visualize the distribution of $Y(t,s)$ for any values of t and s (i.e. as if it was an ergodic process) in Fig. 7. The mean value of this distribution is equal to $\mu = -0.3563$, its standard deviation is equal to $\sigma = 14.6581$.

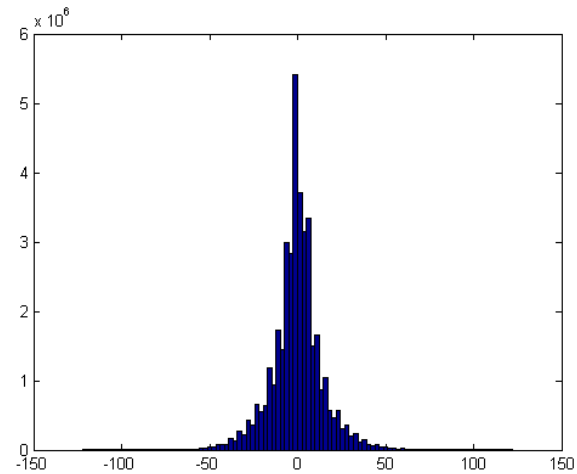


Figure 7: P&S noise distribution.

Characterization of the $Y(t,s)$ distribution. In this experiment, we want to verify whether the $Y(t,s)$ follow a given classical distribution (a normal or a Laplace) or not. In order to answer this question, we compare the distribution of the random process $Y(t,s)$ with each specified distribution using the χ^2 goodness-of-fit test described previously.
A reduced sample data set has been created by randomly selecting $n' = 100$ vectors of $Y(t,s)$ (for lower computational time). We therefore work with a data subset of $9,000,000$ samples. To counteract working with a subsample, we perform each test several times.
An algorithm of χ^2 goodness-of-fit test involving the estimation of mean and variance is presented in Algorithm 1. The result of this test is that both hypotheses are rejected at a significance level

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.8.MWSF-089

of 0.05. Thus our data-set has neither a normal nor a Laplace distribution.
We computed two distances to see how far the empirical CDF based on the initial data is from the CDF of a normal or of a Laplace distribution. This comparison involves the calculation of the maximal distance d_{max} between the CDFs (see the formula (20)) and the squared distance d^2 between the CDFs (see the formula (21)).

Number of samples N	Laplace distibution		Normal distibution	
	d_{max}	d^2	d_{max}	d^2
30,000	0.0405	0.0141	0.0884	0.1208
100,000	0.0471	0.0185	0.0929	0.1297
200,000	0.0434	0.0168	0.0929	0.1247
300,000	0.0431	0.0158	0.0909	0.1225
400,000	0.0424	0.0158	0.0922	0.1233

Table 2: Distances among CDFs of initial data (P&S noise) and estimated Laplace and normal distributions.

Table 2 presents the differences among the CDF based on our data-set and the CDF(s) of a normal and a Laplace distributions. We notice that these distances are rather stable under the changes of sample size.

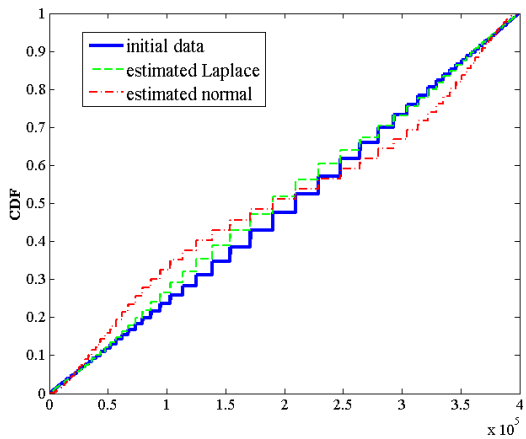
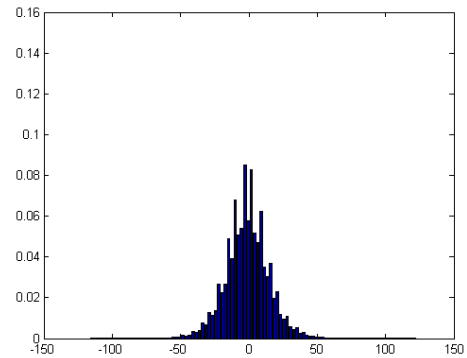


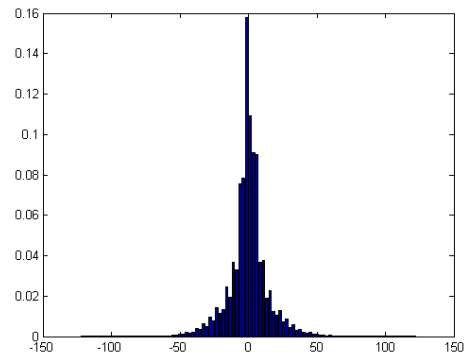
Figure 8: CDF of P&S noise (blue line), estimated Laplace (green line) and estimated normal (red line) distributions.

We illustrate these distances among CDFs in Fig. 8. We note that the CDF of the initial data (blue line) is closer to the CDF of the Laplace distribution (green line) than the CDF of the normal distribution (red line). Therefore the P&S noise can be said to be closer to a Laplace distribution than to a normal distribution.

Additive noise. After a P&S process, the image obtained with one of the textured images consists of gray level pixels when the original textured image is binary. Therefore, all gray level pixels can be separated into two classes: the black class, that consists of black pixels from textured image, and the white class, that consists of white pixels from textured image.
The noise cannot be assumed to be additive if the error distribution of the black pixels is different from the error distribution of the white pixels. As we know the true map of black and white pixel placement, we separate the $\varepsilon_{ii'}$ in two classes: black class and white class. This leads to two random processes $B(t,s)$ and $W(t,s)$. The histograms of error black random processes and white random processes are illustrated in Fig. 9.



(a) Black pixel class



(b) White pixel class

Figure 9: Histograms of P&S noise introduced to a) Black pixels and b) White pixels.

We apply the Mann-Whitney test in order to decide whether or not the error distribution can be said to be additive, i.e. quantify the hypothesis that the black and white distributions are the same. The hypothesis that the distributions are the same is rejected at a significance level of 0.05. Therefore, the noise added by the printing and scanning process cannot be considered as being additive.

Stationarity of first order. To verify the first order stationarity of random process $Y(t,s)$, we use the method described previously. The null hypothesis of Kolmogorov-Smirnov test is formulated as: *the two independent random subsets of mean value set have the same distributions*. The Kolmogorov-Smirnov test cannot reject the null hypothesis at a significance level of 0.05, therefore our process can be supposed being stationary of first order, i.e. the statistical average respects $E[Y(t,s)] = \mu_Y$.

Ergodicity of first order. To verify the first order ergodicity of the random process $Y(t,s)$, we use the method described previously. The null hypothesis is formulated as: *the spatial mean and statistical mean are identical*. The Mann-Whitney test rejects the null hypothesis at a significance level of 0.05. Therefore, our process is not ergodic of first order.

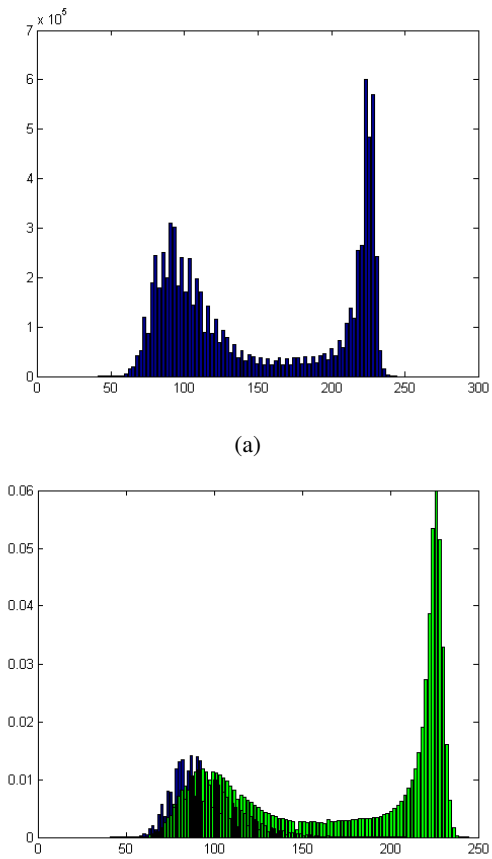
©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.8.MWSF-089

Stationarity of second order. To verify the second order stationarity of the random process $Y(t,s)$, we use the method described previously. The null hypothesis of the Kolmogorov-Smirnov test is formulated as: *the two independent random subsets of autocovariance set have the same distributions*. The Kolmogorov-Smirnov test rejects the null hypothesis at a significance level of 0.05, therefore our process is not stationary of second order. That means our process is not wide-sense stationary.

Test for white noise. A random process can be a white-noise process if it is wide-sense stationary. Our process is not wide-sense stationary, therefore our process is not a white-noise process.

First experiment of color distributions after P&S process

The authors in [1] experimentally show that the black/white block of pixels are related to random variables following asymmetric log-normal distributions. Using our database, we have decided to do the same experiment. Nevertheless, as in our experiments we used textured images, the black and white pixels could be isolated.



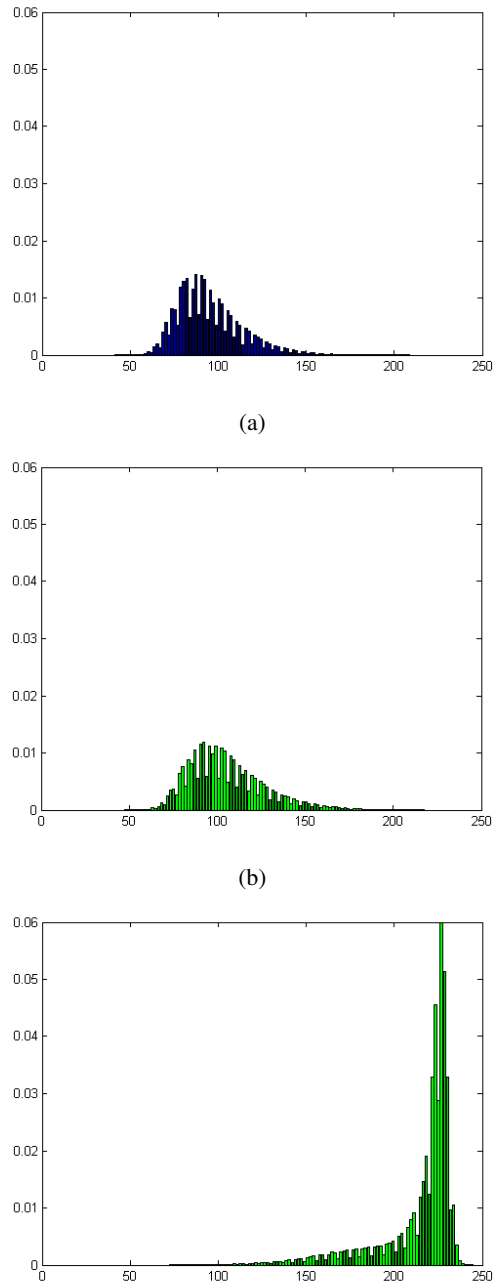
(a)

(b)

Figure 10: Histogram of 90 P&S samples a) all pixels together and b) pixels separated in black (blue color) and white (green color) classes.

The results presented in this section are the work-in-progress, and we suppose to continue the study of color changes after P&S process.

Experimental setup. We print and scan a textured image 90 times. Therefore we have 90 samples of textured images (see Fig.2). The histogram of all these images is illustrated in Fig. 10.a. Then, using the true map, we separate our pixels in two classes: black class and white class. We illustrated the histograms of obtained classes in Fig. 10.b.



(a)

(b)

(c)

Figure 11: Histogram of 90 P&S samples a) Black pixels, b) White pixels from textured patterns and c) White pixels from white modules.

DOI: 10.2352/ISSN.2470-1173.2016.8.MWSF-089

As expected, the histogram of black pixels is uni-modal (see Fig. 11.a). However, the bi-modal histogram of white pixels needs a more throughout study. Indeed, our P&S samples contain three different kinds of pixels: black pixels from textured patterns, white pixels from textured patterns and white pixels from completely white modules (see Fig. 2). Therefore, we separate the white pixels again in two classes: white pixels of textured patterns and white pixels of white modules. The results of this separation are illustrated in Fig. 11.b and Fig. 11.c.

These histograms (Fig. 11.b and Fig. 11.c) clearly show that the white pixels of textured patterns are placed at the same part of histogram as the black pixels. This could be simply explained by the structure of our textured patterns, where the black and white pixels change frequently.

We have applied the χ^2 goodness-of-fit test to verify whether these distributions have normal or log-normal nature. The χ^2 test rejects the null hypothesis, that these distributions are normal, at a significance level of 0.05. In the same time, it accepts the null hypothesis, that they are log-normal, at a significance level of 0.05.

These results are only the beginning of this study. These experiments need to be evaluated better, using different statistical tests. These first results highlight several open questions:

- The histograms seem to present artifacts (looking like quantification effects). We should find the nature of these artifacts and find the possibility to remove them.
- Raw histograms seem to present a log-normal distribution. Several additional statistical test need to be done in order to confirm or disprove these observations.

Conclusions

In this paper we proposed to experimentally study the impact of P&S process as a random process. The main results of this paper are: the characterization of the scanning noise is done independently from the printing noise, the invalidation of the classical hypothesis that the P&S noise is normal distributed, and either white or ergodic. Additionally, we show that the scanning noise and the P&S noise are both closer to Laplace distribution, than to normal distribution. The first experiment of color distributions after P&S process are presented. We conclude that this problem needs a more throughout study, but the first experimental results are interesting.

In future work we want to study the quantification effects presented in histograms, try to find the parameters of P&S pixel distribution, and simulate this P&S process using convolution of original textured image with a filter extrapolated from the P&S distribution.

References

[1] C. Baras and F. Cayre, Towards a realistic channel model for security analysis of authentication using graphical codes, Information Forensics and Security (WIFS), 2013 IEEE International Workshop on, p. 115-119, 2013.

[2] L. Yu, X. Niu and S. Sun, Print-and-scan model and the watermarking countermeasure, Image and Vision Computing, Vol. 23, Num. 9, p. 807-814, 2005, Elsevier.

[3] O. Ibe, Fundamentals of applied probability and random processes, Academic Press, 2014.

[4] Iu. Tkachenko, W. Puech, O. Strauss, J.-M. Gaudin, C. Destruel, and C. Guichard, Fighting Against Forged Documents by Using Textured Image, Signal Processing Conference (EUSIPCO), Proceedings of the 22th European, 2014.

[5] A. Vongkunghae, J. Yi and R. B. Wells, A printer model using signal processing techniques, Image Processing, IEEE Transactions on, Vol. 12, Num. 7, p. 776-783, 2003.

[6] J. Tchan, RC. Thompson and A. Manning, A computational model of print-quality perception, Expert Systems with Applications, Vol. 17, Num. 4, p. 243-256, Elsevier, 1999.

[7] C-Y. Lin and S-F. Chang, Distortion modeling and invariant extraction for digital image print-and-scan process, Int. Symp. Multimedia Information Processing, 1999.

[8] K. Solanki, U. Madhow, B. S. Manjunath and S. Chandrasekaran, Modeling the print-scan process for resilient data hiding, Electronic Imaging, International Society for Optics and Photonics, p. 418-429, 2005.

[9] T. D. Kite, B. L. Evans and A. C. Bovik, Modeling and quality assessment of halftoning by error diffusion, Image Processing, IEEE Transactions on, Vol. 9, Num. 5, p. 909-922, 2000.

[10] S. Voloshynovskiy, O. Koval, F. Deguillaume and T. Pun, Visual communications with side information via distributed printing channels: extended multimedia and security perspectives, Electronic Imaging, International Society for Optics and Photonics, p. 428-445, 2004.

[11] A-T. Phan Ho, B-A. Mai Hoang, W. Sawaya and P. Bas, Authentication using graphical codes: Optimisation of the print and scan channels, Signal Processing Conference (EUSIPCO), Proceedings of the 22nd European, p. 800-804, 2014.

[12] A-T. Phan Ho, B-A. Mai Hoang, W. Sawaya and P. Bas, Document Authentication Using Graphical Codes: Reliable Performance Analysis and Channel Optimization, EURASIP Journal on Information Security, Springer, Vol. 2014/1/9, 2014.

[13] R. R. Wilcox, Introduction to robust estimation and hypothesis testing, Academic Press, 2012.

[14] D. J. Sheskin, Handbook of parametric and nonparametric statistical procedures, CRC Press, 2003.

Author Biography

Iuliia TKACHENKO *received the M.S. degree in Applied Mathematics from Dnipropetrovsk National University, Ukraine in 2010, the M.S. degree in Cryptography and Information Security from University Bordeaux 1, France in 2012 and the Ph.D. degree in Computer Science from the University of Montpellier, France, in 2015. Currently she is an assistant researcher in Authentication Industries and collaborates with the Montpellier Laboratory of Informatics, Robotics and Microelectronics (LIRMM), France. Her research interests include multimedia security, document authentication, information hiding and cryptography.*

William PUECH *received the diploma of Electrical Engineering from the University of Montpellier, France, in 1991 and the Ph.D. Degree in Signal-Image-Speech from the Polytechnic National Institute of Grenoble, France in 1997. He started his research activities in image processing and computer vision. He served as a Visiting Research Associate to the University of Thessaloniki, Greece. From 1997 to 2000, he had been an Assistant Professor in the University of Toulon, France, with research interests including methods of active contours applied to medical images sequences. Between 2000 and 2008, he had been Associate Professor and since 2009, he is full Professor in image processing at the University of Montpellier, France. He works in the LIRMM Laboratory (Laboratory*

DOI: 10.2352/ISSN.2470-1173.2016.8.MWSF-089

of Computer Science, Robotic and Microelectronic of Montpellier). His current interests are in the areas of protection of visual data (image, video and 3D object) for safe transfer by combining watermarking, data hiding, compression and cryptography. He has applications on medical images, cultural heritage and video surveillance. He is the head of the ICAR team (Image & Interaction) and he has published more than 15 journal papers, 8 book chapters and more than 80 conference papers. W. Puech is associate editor of J. of Advances in Signal Processing, Springer, Signal Processing: Image Communications, Elsevier and Signal Processing, Elsevier and he is reviewer for more than 15 journals (IEEE Trans. on Image Processing, IEEE Trans. on Multimedia, IEEE TCSVT, IEEE TIFS, Signal Processing: Image Communication, Multimedia Tools and Applications ...) and for more than 10 conferences (IEEE ICIP, EUSIPCO, ...).

Olivier STRAUSS *is an associate professor in Signal and Image Processing in Montpellier University, France. He received his Ph.D. in Signal Processing from the Montpellier University in 1991 and obtained his accreditation to supervise research in 2008. His current research interests include signal and image processing, computer vision, medical imaging, uncertainty management, advanced statistics and imprecise probability theories. He is currently head of the robotics department and member of the scientific council of Montpellier University.*

Christophe DESTRUEL *received his Engineer's Degree from the University of Toulouse, France, in 1995. He worked for 10 years in collaboration with the french Spatial Agency on research programs related to its main fields of interest, Computer Graphics and Image Processing, to develop the use of 3D paradigm in space imaging. He is now the Scientific Director of the start-up Authentication Industries that proposes innovative solutions to authenticate valuable documents. Respecting industrial constraints, his contributions have been published in various international conferences.*