

Image Forensics of Digital Cameras by Analysing Image Variations using Statistical Process Control

Philip Bateman
and Anthony T. S. Ho
Department of Computing
University of Surrey
Guildford, GU2 7XH UK.

Email: P.Bateman@surrey.ac.uk, A.Ho@surrey.ac.uk

Alan Woodward
Charteris Plc.
Charteris House
39/40 Bartholomew Close
London, EC1A 7JN UK.

Email: Alan.Woodward@charteris.com

Abstract—In this paper, we propose the novel use of Statistical Process Control (SPC) as a tool for identifying anomalies in digital cameras, by analysing image variations. We apply SPC on six low-end cameras, and one mid-range camera. By plotting the mean distribution of image pixel data through control charts, we have been able to identify a clear distinction between images captured from low-end cameras (where the variation is approximately 21%), and mid-range cameras (where the variation is approximately 1%). Control charts are used to highlight *out-of-control* values within the image data. By examining the data of any out-of-control images, it is possible to deduce the cause of inconsistency in the device's image acquisition process. This could ultimately lead to the identification of a stochastic feature for images taken by certain types of digital cameras, to benefit research for camera identification.

In this paper, the SPC model is described in relation to its potential use in the field of image forensics for aiding camera identification, and some experimental results are shown from the six test cameras to demonstrate the effectiveness of the model.

I. INTRODUCTION

Whilst research in the field of image forensics is very much in its infancy, the importance of reliable schemes capable of correctly identifying the digital camera that captured a particular image is immediately apparent. Perhaps the most significant potential application for such schemes is the usage in the court for proving the origin of unlawful images, which would go some way to helping witnesses to identify where the blame lies. Similarly, the schemes could also prove useful for proof of ownership and subsequently copyright infringement cases.

Recently, much research has been directed towards identifying components of digital cameras that create anomalies in the image data, thereby presenting a unique "fingerprint" that can ultimately be used to map the image to the source camera. Several methods for identifying unique characteristics of cameras through the corresponding image data have been proposed in recent years. In [1], the sensor is shown to produce noise patterns that invoke a unique signature. In [2], Choi et al. proved that radial distortion (originating from the camera lens, causing straight lines to appear curved), are somewhat different for each make of camera. As each make of camera possesses a unique radial distortion pattern, the pattern can be

used to determine whether an image originated from a suspect camera. In addition to these methods, research against Colour Filter Array (CFA) interpolation has also been carried out in [3], [4], and [5], as well as a feature extraction scheme in [6] that extracts 34 specially selected image features in order to uniquely classify a camera model.

The above-mentioned camera identification methods are all made possible by the fact that each digital camera manufacturer is likely to use different internal components that vary in cost and quality. As a result, a sample of images captured from a range of different digital cameras is likely to appear different in some way, even if the images are of the exact same scene. The main challenge lies in isolating the components that cause the biggest shift in variation in terms of the image data. These include internal camera components such as the lens, filters, sensors, A/D conversion process, denoising, white balancing, as well as external influences such as lighting, shadowing, and temperature.

At present, there is no reported method for identifying the most suitable feature of a camera's image acquisition process that will be a good "fingerprint" for the device. In this paper, we propose the novel application of statistical process control (SPC) as an image forensics technique to highlight inconsistencies in the image data, which can help to make such an identification.

II. STATISTICAL PROCESS CONTROL

A. Overview

The theory of SPC was developed in the late 1920s by Dr. Walter Shewhart, a statistician at the AT&T Bell Laboratories, USA. Shewhart observed that repeated measurements taken from a process will exhibit variation that can be classified as either *common-cause* or *special-cause* [7]. Common-cause variation refers to natural variation that the process is expected to inherit, and special-cause variation refers to unnatural variation caused by irregular events or circumstances that have an ultimate impact on the process. In an ideal world, each measurement will produce results that are nearly identical, indicating that the process is in good control. However, for a vast quantity of real-life applications (for example, the time it

takes to spray paint a vehicle on the production line), the results are likely to differ slightly for each measurement. Occasionally, an event might occur that pushes the measurement outside the expected range, consequently producing special-cause variation. In manufacture, processes containing such events can be improved by identifying and removing the special-cause to bring the process into better control.

The use of SPC has mostly been geared towards quality control in the manufacturing industry with the aim to maximise the efficiency of production processes in order to deliver high quality products or services. SPC was first applied to automobile manufacture by several Japanese manufacturers, and was so prominent in producing high quality products that the Ford Motor Company soon followed [8]. It has also been applied for a range of industrial applications, including the pulp and paper industry [9]-[11], and has more recently been considered as a tool for research and healthcare improvement [12].

Whilst the image acquisition process for digital cameras comprises many components and sub-processes, our aim is to treat every element as a *black-box* process, such that we evaluate the complete unit at once. SPC can be used to derive the complete depth of variance for a range of cameras under the same environmental conditions. It is hypothesised that low-end consumer cameras will output a wider depth in variation as compared to high end cameras as the components and correction algorithms are much less refined.

B. Control Charts

Control charts are a key tool of SPC that help to distinguish between common-cause and special-cause variation for a particular process [7]. Typically, two control charts are used to graphically display the variation obtained for a given process: one to display the shifts in the process mean, and one to display the shifts or changes in the amount of process variability. There are many types of control charts, and it depends on how the process is monitored and how the results are obtained, as to which chart should be used. Throughout this paper we use an X chart (often referred to as an *Individuals* chart) to display the shifts in the process mean, and an R_m chart (referred to as a *Moving Range* chart) to plot the process variability, as these charts are the most suited to analysing data taken from individual samples.

Control charts comprise a *centreline* (CL) (the mean), an *upper control limit* (UCL), and a *lower control limit* (LCL), collectively referred to as *control limits*. Shewhart proposed that the UCL and LCL control limits are calculated as ± 3 standard deviations of the CL in order to detect meaningful changes in performance, with a low false-positive margin of approximately 0.27% [12]. Any sample measurements falling above or below these limits corresponds to a value that is *out-of-control*.

C. Calculating the Control Limits

The R_m chart is generally plotted first as it illustrates the overall process variability. The first step to creating an R_m

chart is to determine the difference between the neighbouring sample values (referred to as X). The CL is calculated by determining the average of all samples in R_m as shown in Equation (1).

$$\bar{R}_m = \frac{\sum_{i=1}^k R_{mi}}{k}. \quad (1)$$

where k refers to the total number of elements in R_m . The UCL and LCL are calculated as given in Equation (2).

$$\begin{aligned} UCL_{R_m} &= D_4 \bar{R}_m \\ LCL_{R_m} &= D_3 \bar{R}_m. \end{aligned} \quad (2)$$

where D_3 and D_4 are constants obtained from a table of constants as shown in Table I, where *observations in sample* $n = 2$.

TABLE I
CONSTANTS FOR CALCULATING CONTROL LIMITS [13].

Observations in Sample	d_2	A_2	D_3	D_4
2	1.128	1.880	0	3.267
3	1.693	1.023	0	2.575
4	2.059	0.729	0	2.282
5	2.326	0.577	0	2.115
6	2.534	0.483	0	2.004
7	2.704	0.419	0.076	1.924
8	2.847	0.373	0.136	1.864
9	2.970	0.337	0.184	1.816
10	3.078	0.308	0.223	1.777
15	3.472	0.223	0.348	1.652
20	3.735	0.180	0.414	1.586

Similarly, when dealing with X charts, CL refers to the grand average value \bar{X} of all the sample averages in X . The UCL and LCL values are calculated in accordance with \bar{X} as well as an estimate of the standard deviation of X , denoted by $\hat{\sigma}$. An estimate of the standard deviation is required for determining whether or not the data values are *controlled* or *out-of-control*, and is calculated using Equation (3).

$$\hat{\sigma}_X = \frac{\bar{R}_m}{d_2}. \quad (3)$$

where the actual value for d_2 is taken from the table of constants where observations in sample $n = 2$.

The UCL and LCL control limits are calculated by adding or subtracting 3 standard deviations respectively.

$$\begin{aligned} UCL_X &= \bar{X} + 3\hat{\sigma}_X \\ LCL_X &= \bar{X} - 3\hat{\sigma}_X. \end{aligned} \quad (4)$$

Once plotted, the control charts clearly indicate how closely the sample data is centred around the process mean, and therefore how controlled the process is.

III. DATA ACQUISITION

A. Experiment

In our experiments, SPC is used to evaluate the complete process of a digital camera's image acquisition process by collecting 10 consecutive images of an identical scene from a range of cameras. Our primary test devices are four Apple iPhone 3G models. At the time of writing, Flickr (the popular image and video hosting site) states that this device is the current most frequently used camera model by its users. The iPhone 3G is not primarily engineered for photography, and as such uses inexpensive camera components. The image resolution is 2 MegaPixels (1600x1200), and there exists no optical or digital zoom. There are also no settings that can be changed, meaning the error correction and image enhancement processes are automatic (if they even exist at all). It is therefore expected that SPC will show the image acquisition process to be minimally controlled.

A Sony Ericsson W810i camera phone was also obtained for comparison with the iPhone 3G devices. The W810i also comprises a 2 MegaPixel sensor. Finally, a standalone digital camera, a Samsung NV3, is used for comparison with the cameraphones. To ensure a fair test, all of the settings on both of these devices were set to match those of the Apple iPhone 3G (the most basic camera), so that the inferred performance of each device is more appropriately comparable.

B. Test Environment

It is very important that the environment in which the images are captured is controlled as much as possible to ensure that it has a restricted impact on the image acquisition process. Without controlling the environment, it might be possible that each device produces different results, not because of the camera make-up, but because of an external factor such as temperature changes, or changes in ambient lighting due to sunlight or cloud cover. In order to nullify the external influences, we took the images from inside a room that had no windows, meaning that the ambient lighting is more controlled. The scene is lit via fluorescent lighting, which is well known for flickering. Whilst this is not ideal for most experiments, it is quite useful for testing how effectively each device reacts to the light flicker. The room was also air conditioned to a constant temperature to reduce the potential impact of temperature changes affecting the results.

C. Test Scene

The scene itself comprises a white bowl containing colourful confectioneries that force the images to inherit a wide variation of colour shifts. Whilst colour shifts of this degree are rarely seen in real world scenes, it acts as a good method for testing the colour interpretation for each camera.

Surrounding the bowl, are 10 angle reference points which are used to align each device to the same position. When introducing angles, we simultaneously introduce natural shadowing from the object matter, meaning we can later review whether or not the discrepancies are connected to lighting, if need be. However, what we are mainly reviewing is the image data

obtained per device, on a per angle basis. Comparisons are made across all iPhone data solely to confirm that each model inherits the same characteristics. Ten images are captured from each of these ten angles, meaning 100 images are taken for each device. A bendable tripod is used to ensure that the images taken from all devices are as similar as possible.

IV. RESULTS

Having captured all the images from each device, the mean pixel value was calculated for each image. The mean value is calculated across all colour planes in the spatial domain, and provides information as to how diverse the image data is for all the cameras at each angle. Table II shows the mean pixel values obtained from each iPhone where angle $\angle = 5$.

TABLE II
MEAN PIXEL VALUES OBTAINED FOR EACH APPLE IPHONE 3G.

Shot No.	iPhone A	iPhone B	iPhone C	iPhone D
1	110.2097	105.4888	104.6182	104.4518
2	109.6045	106.5222	104.6503	97.4483
3	109.1334	106.7678	105.4275	97.1251
4	109.1161	98.0294	105.4614	97.0542
5	109.2108	98.064	105.676	97.0346
6	101.3616	97.7303	105.1278	97.0085
7	101.7246	98.9264	105.4571	97.4346
8	101.2875	96.9707	105.5588	97.4245
9	101.2724	98.4508	105.5459	97.0113
10	101.6922	97.4999	105.4956	97.0198

The captured images vary quite significantly from shot to shot, particularly for iPhones A, B, and D. Whilst iPhone C does not display the same inconsistencies at this angle, it was inconsistent for others. When $\angle = 2$, for example, the data for iPhone C ranges from 108.9369 through 121.0634, whilst iPhones B and D are more finely controlled. \bar{X} and R_m control charts are constructed for every device based on the data presented in Table II.

Using iPhone A as an example, the R_m values are calculated first. The results of which are shown in Table III.

TABLE III
MOVING RANGE VALUES OBTAINED FROM IPHONE A DATA.

Shot No.	\bar{X}	R_m
1	110.2097	
2	109.6045	0.605
3	109.1334	0.471
4	109.1161	0.017
5	109.2108	0.095
6	101.3616	7.849
7	101.7246	0.363
8	101.2875	0.437
9	101.2724	0.015
10	101.6922	0.420

The CL value is calculated as 1.141. Using the table of constants from Table I, where the sample $n = 2$, we obtain $D_3 = 0$ and $D_4 = 3.27$. The UCL and LCL are calculated according to Equation (5).

$$\begin{aligned}
UCL_{R_m} &= D_4 \bar{R}_m = (3.27)(1.141) = 3.729 \\
LCL_{R_m} &= D_3 \bar{R}_m = (0.0)(0.294) = 0.0
\end{aligned} \quad (5)$$

The centerline for the X chart is calculated as 105.461. Equation (3) is used to obtain the standard deviation, which is calculated as 1.011.

The X -chart control limits are calculated by adding or subtracting 3 standard deviations from the process mean, as shown in Equation (6).

$$\begin{aligned}
UCL_X &= \bar{X} + 3\hat{\sigma}_X = 105.461 + (3)(1.011) = 108.497 \\
LCL_X &= \bar{X} - 3\hat{\sigma}_X = 105.461 - (3)(1.011) = 102.426.
\end{aligned} \quad (6)$$

The X and R_m charts are calculated in the same way for each device. Fig. 1 shows the control charts for iPhone A. The circled data points denote that the measurement is *out-of-control* because it falls either above or below the control limits.

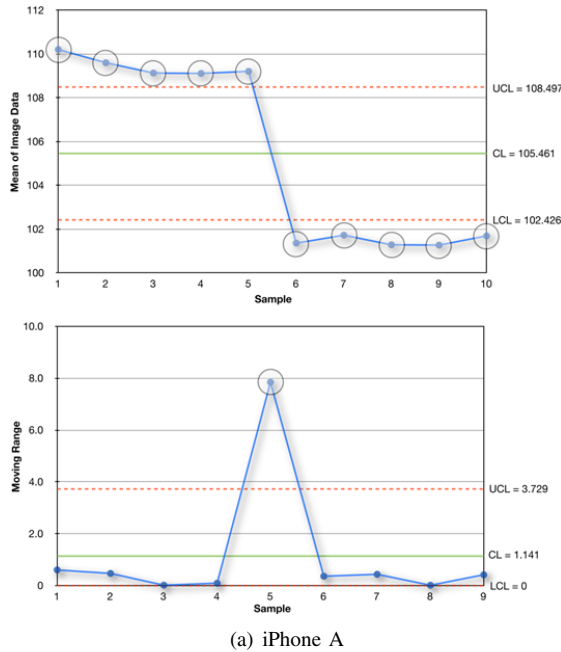


Fig. 1. X Charts and R_m Charts for iPhone.

The control charts show that every data value for this device is out-of-control. This indicates that the overall quality of the image acquisition process is unstable, at least for the shots taken at this angle. The R_m chart indicates that the error seems to fall around the fifth measurement. On the X chart, this corresponds to images 5 and 6, where we witness a quite significant drop in values. By reviewing these two images in greater detail, we can see that there is a significant change in brightness between these two images.

The control charts are similar for iPhone B at the same angle, and a wide shift in variation is observed. Again, there

are many out-of-control values indicating that the process is not statistically stable. When evaluating the images that are significantly varied, it could again be noted that the brightness is significantly different.

For comparison, the same tests were performed on the Sony Ericsson W810i and Samsung NV3 devices, where $\angle = 5$. We would expect the quality of the image acquisition process for the Samsung NV3 camera to be better than that of the Apple iPhone 3G and the Sony Ericsson W810i, simply on the basis that it is a standalone digital camera. The X and R_m data for both devices are shown in Table IV.

TABLE IV
MOVING RANGE VALUES OBTAINED FROM SONY ERICSSON W810I AND SAMSUNG NV3.

Shot No.	Sony Ericsson W810i		Samsung NV3	
	X	R_m	X	R_m
1	127.6547		132.0012	
2	126.5369	1.118	131.8909	0.110
3	126.7548	0.218	131.9225	0.032
4	128.0077	1.253	131.9527	0.030
5	127.9382	0.070	131.9752	0.023
6	128.1914	0.253	131.955	0.020
7	128.0699	0.121	131.996	0.041
8	126.7867	1.283	131.9873	0.009
9	126.7496	0.037	132.0464	0.059
10	127.9688	1.219	132.0355	0.011

By plotting the X and R_m control charts for the Sony Ericsson W810i (as shown in Fig. 2), we can clearly see that the process is more stable than that for the iPhones. For the W810i, there are no out-of-control data points, meaning the image acquisition process is in good statistical control. The highest value X chart for this angle ($\angle = 5$) is 132.046, whilst the lowest value is 131.891. The difference between the highest and lowest points is 0.155. Whilst it is not possible to show the results for each angle, the results are extremely similar in every case.

The control charts for the Samsung NV3 imply a further improvement, as shown in Fig. 3. Again, both control charts are in good statistical control, with the difference between the highest and lowest points (132.04 and 131.89, respectively) reduced to 0.15. Similar difference rates are obtained from the data recorded at each angle.

In order to highlight the depth of variation for each device, we calculate the percentage shift for each angle, and plot the lowest and highest scores. Fig. 4 shows that the iPhones each inherit the widest variation by some margin; the worst case being iPhone C with a variation range of approximately 21%. The depth of variation for the Sony Ericsson W810i and the Samsung NV3 cameras is much smaller (2% and 1% respectively), indicating that the image acquisition process is much more controlled.

V. CONCLUSION

It was not our expectation that the iPhone 3G cameras would perform so badly in comparison to the Sony Ericsson W810i camera (21% vs. 1% variation), but by adopting an

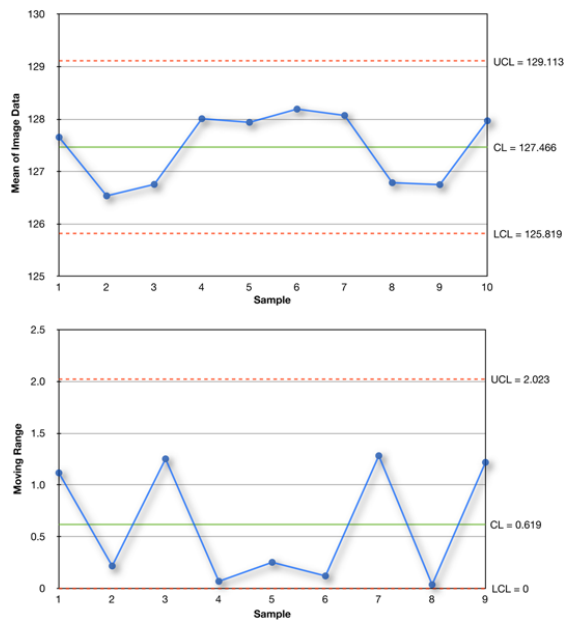


Fig. 2. \bar{X} Chart and R_m Chart for Sony Ericsson W810i.

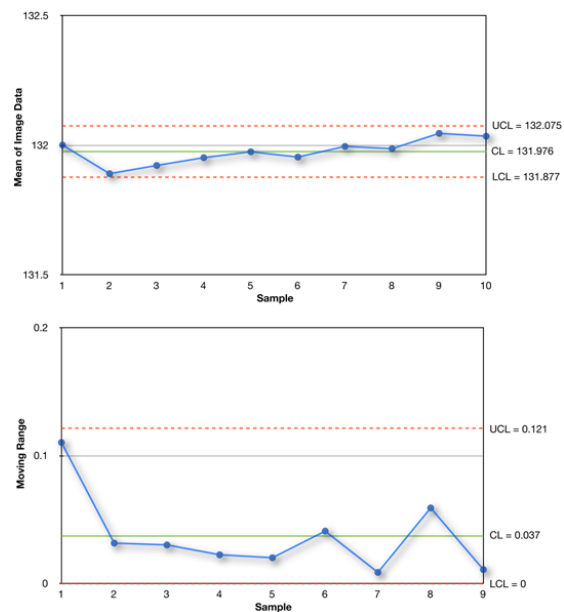


Fig. 3. \bar{X} Chart and R_m Chart for Samsung NV3.

SPC model, our experimental results have shown that this is the case. We also observe that some component (or lack of component) in the iPhone image acquisition process is causing the brightness of images to vary quite significantly on certain shots. It is likely that a scrutinised analysis of the image data would show that the iPhone 3G possesses exposure problems, which could easily help to link future images to these devices.

The use of SPC has been shown to act as a tool for yielding anomalies in image data. When applied to digital images, SPC techniques make it possible to pinpoint inconsistent images such that it is potentially possible to establish a stochastic

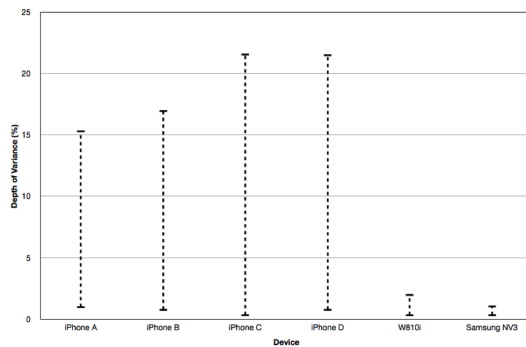


Fig. 4. Depth of variation for all devices.

feature in the image data that acts as a fingerprint for that device. Our future work will focus on the application of *Cause & Effect* diagrams and *Pareto* charts to determine whether they can be used to further aid the identification of special-causes, based on pattern analysis of out-of-control points. This research could also lead to the identification of a stochastic feature for a range of devices that are of the same make or model.

ACKNOWLEDGMENT

This research is funded by an EPSRC CASE award, in collaboration with Charteris Plc.

REFERENCES

- [1] J. Lukáš, J. Fridrich, and M. Goljan, "Digital Camera Identification From Sensor Pattern Noise," *IEEE Transactions on Information Security and Forensics*, vol. 1(2), pp. 205-214, 2006.
- [2] K. S. Choi, E. Y. Lam, and K. K. Y. Wong, "Source Camera Identification Using Footprints From Lens Aberration," *Proceedings of the SPIE*, vol. 6069, pp. 172-179, 2006.
- [3] S. Bayram, H. T. Sencar, N. Memon, and I. Avcibas, "Source Camera Identification Based on CFA Interpolation," *Proceedings of IEEE ICIP*, vol. 3, pp. 69-72, 2005.
- [4] O. Celiktutan, I. Avcibas, B. Sankur, and N. Memon, "Source Cell-phone Identification," *IEEE Signal Processing and Communications Applications*, pp. 1-3, 2005.
- [5] Y. Long and Y. Huang, "Image Based Source Camera Identification using Demosaicking," *IEEE 8th Workshop on Multimedia Signal Processing*, pp. 419-424, 2006.
- [6] M. Kharrazi, H. T. Sencar, and N. Memon, "Blind Source Camera Identification," *International Conference on Image Processing*, vol. 1, pp. 709-712, 2004.
- [7] W. A. Shewhart, "Economic Control of Quality of Manufactured Product," *American Society for Quality*, ISBN: 978-0873890762, 1931.
- [8] Ford Motor Company, "Continuing Process Control and Process Capability Improvement," *Manual Published by Statistical Methods Office*, 1984.
- [9] A. T. S. Ho, and C. Henriksson, "Improving Product Quality in a Pulp Mill Using Statistical Process Control (SPC)," *IEEE Canadian Conference on Electrical and Computer Engineering*, pp. 1139-1143, 1993.
- [10] A. T. S. Ho, and C. Henriksson, "Improvement of Product Uniformity Using Statistical Process Control (SPC) in a BCTMP Mill," *Pulp and Paper Canada*, vol. 95, pp. 37-40, 1994.
- [11] A. L. Guillory, "Statistical Process Control in a Paper Mill," *Chemical Engineering Progress*, vol. 84, pp. 52-57, 1988.
- [12] J. C. Benneyan, R. C. Lloyd, and P. E. Plsek, "Statistical Process Control as a Tool for Research and Healthcare Improvement," *Qual Saf Health Care*, vol. 12, pp. 458-464, 2003.
- [13] R. E. DeVor, T. Chang, and J. Sutherland, "Statistical Quality Design and Control: Contemporary Concepts and Methods," *Prentice Hall*, ISBN: 978-0023291807, 1992.