

Accurate Detection of Out-of-Control Variations From Digital Camera Devices

Philip Bateman, Anthony T.S. Ho
Department of Computing
University of Surrey
Guildford, UK.
{P.Bateman, A.Ho}@surrey.ac.uk

Alan Woodward
Charteris Plc.
39/40 Bartholomew Close
London, UK.
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 the image acquisition process of a digital camera, for the purpose of camera identification. Control charts are used to illustrate the overall level of control associated with several devices (models include Apple iPhone 3G and 3GS, Nokia N97, and Leica D-Lux4), which are in turn reviewed in accordance with the Western Electric Rules for identifying assignable causes for the observed variation. X-Moving Range and Exponentially Weighted Moving Average (EWMA) control charts are used to highlight the variation for a subset of the devices. By implementing such a statistical model, the forensic investigator is much better positioned to understand the behaviour of a particular device, and is ultimately able to identify the most unstable feature of the cameras image acquisition process, thereby establishing a suitable fingerprint for matching images to their source.

I. INTRODUCTION

From a legal perspective, there exist several scenarios where it would be useful to successfully prove the origin of an image. If illegal images are found on a suspect's computer, it would be useful to ascertain whether or not those images were also captured by the suspect. Recent research in the field of camera identification has led to the identification of key components in the image acquisition process that cause the output image data to be deterministic on some level. The image sensor is described in [1] as yielding a unique noise pattern for each device, and as this noise pattern remains the same for all images captured by the same device, it is useful for linking images to the source. Similarly, radial distortion has also been found to differ vastly for a range of devices, and also makes for a useful fingerprint [2]. In addition, a feature extraction scheme is described in [3] that extracts 34 intrinsic features of an image, and uses an SVM to classify the most likely source device.

In this paper, we present an alternative and novel approach to camera identification, by modelling the individual behaviour of popular cameraphone devices (including 2 Apple iPhone 3G's, an iPhone 3GS, 2 Nokia N97's, and a Leica D-Lux 4 standalone camera), using Statistical Process Control. As each device monitored is engineered to different standards by the manufacturer, the image acquisition process of each is likely to differ. Furthermore, the overall quality of the components and processes associated with the image acquisition process for a

particular device will cause the image content to vary slightly on a shot-to-shot basis. By reviewing the complete process as a black-box, a forensic expert can model how the process is affected by known external environmental conditions to ultimately learn the behaviour of any device. If we can understand how a device performs, its behaviour can be predicted, meaning if a suspicious image is presented, the forensic expert is then well placed to confidently associate it with the most likely source camera.

We build upon our work in [4] by controlling the test environment such that the image acquisition process is minimally affected by external influences. This is important for ensuring that the witnessed variation is an accurate record of the image acquisition process over time. The control charts (used to illustrate the level of control for each device) are reviewed in accordance with the highly acclaimed Western Electric Rules printed in 1956 [5]. Finally, the Shewhart X-Moving Range control charts are compared against Exponentially Weighted Moving Average (EWMA) charts in an effort to establish the most suitable implementation from a forensic standpoint.

II. STATISTICAL PROCESS CONTROL

Statistical Process Control (SPC) is typically applied for measuring the quality of certain processes in the manufacturing industry [6]-[9]. The theory was developed by statistician Walter Shewhart in the 1920's, as a result of the observation that processes incorporate variation if measurements are repeatedly taken over time [10]. Often, the cause for variation can be explained, and is actually expected (referred to as *common-cause* variation), but sometimes an unforeseen event occurs which has a detrimental effect on the output of the process (referred to as *special-cause* variation). SPC models highlight both common-cause and special-cause variation through *control charts*, and by reviewing these charts, the assignable causes can be dealt with appropriately to improve the overall quality of the process.

Control charts illustrate the variation associated with a particular process, and are presented in [4] for use with digital images. A control chart consists of 4 properties: the image data (calculated as the mean pixel value across each RGB colour plane for each consecutive image); a centreline *CL* (the mean obtained from all data values); an Upper Control Limit

UCL (calculated as 3 standard deviations above *CL*); and a Lower Control Limit *LCL* (calculated as 3 standard deviations below *CL*). When setting the *UCL* and *LCL* to ± 3 standard deviations of the centreline, it means that any value falling inside of these limits is in statistical control with a false-positive margin of approximately 0.27% [12]. Values falling outside of these limits are therefore regarded as *out-of-control*; an indication that the image acquisition process is flawed.

By capturing 10 consecutive images from several devices, we observed that SPC reacts accordingly when applied to devices of varying qualities. Cameraphones, for example, inherit a much wider degree of variation as compared with standalone digital cameras. Furthermore, it was noted that the control charts not only illustrate the variation, they also allow for an identification of images that are statistically *out-of-control*. A review of these images captured from four separate iPhone 3G devices showed that they all inherited the same property whereby the brightness of the image would dramatically increase or decrease at seemingly random intervals.

A. Improving the Test Parameters

The environmental conditions from [4] were modified to improve our confidence that the observed variation is mostly a reflection of the device, and not a result of changing environmental conditions. The light source was replaced with a flicker-free SoLux task lamp, that outputs a D50 colour spectrum, closely matching natural daylight. The subject of the photographs was also changed to an *x-rite ColorChecker*. This chart is comprised of colours associated with natural environments (i.e. skin tones, foliage, etc.), which are reflected the same irrespective of the light source. This provides more usefulness to the captured images, as each test image can be compared to monitor the behaviour of colour interpretation for a specific devices. Each image was captured through a photographic light tent to diffuse any unwanted external light, and to further control the light source. Finally, the number of samples was also increased to 30 in an attempt to create a wider appreciation of the behaviour of each device.

Figure 1 depicts the control chart obtained from an Apple iPhone 3G device under the new controlled environment. The control chart clearly illustrates that the data is out-of-control for the most part, and once again highlights a distinct drop in value between images 17 and 18 where the brightness is dropped significantly. The control chart is similar to that obtained from the initial experiment in [4], inferring that the environment had a minimal impact on the image data taken from each sample. Indeed, similar control charts are produced for each device (four iPhone 3G cameraphones, one Sony Ericsson W810i cameraphone, and one Samsung NV3 standalone digital camera), which further confirms this point.

III. CONTROL CHART INTERPRETATION

Referring back to the usage of control charts in manufacture, a set of decision rules was developed by the Western Electric Company (published in the Quality Control Handbook in 1956) for interpreting trends in the data [5]. The rules made

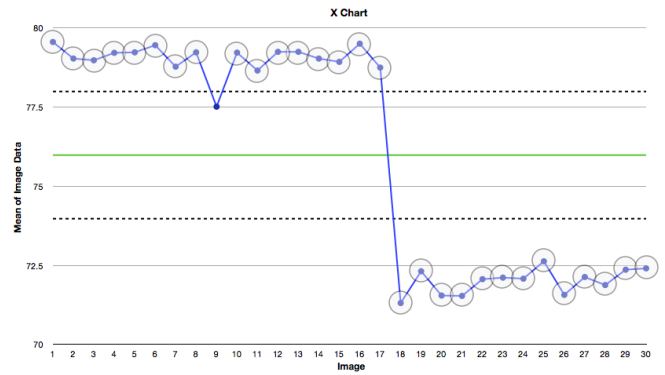


Fig. 1. X Chart for iPhone 3G.

it possible to distinguish between common-cause and special-cause variation.

In order to interpret the control charts, the rules require that the charts are divided into zones. Table I defines each zone in terms of units of standard deviation from the centreline.

TABLE I

CHART ZONES AS DEFINED BY WESTERN ELECTRIC [5].

Zone	Region
Zone A	Within 2σ of the centreline and the control limit
Zone B	Within 1σ and 2σ of the centreline
Zone C	Within 1σ of the centreline

To illustrate these zones by example, Figure 2 shows the control chart for a different iPhone 3G device, with the aforementioned zones highlighted.

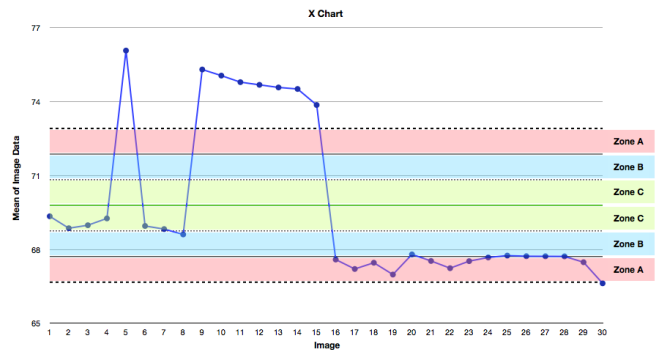


Fig. 2. X Chart for iPhone 3G with highlighted zones.

The Western Electric Company described rules associated with these zones that would be capable of detecting instability within a process. The four most significant rules are:

- 1) Any single data point falls beyond zone A.
- 2) Two out of three consecutive points fall in zone A or beyond, on the same side of the centreline.
- 3) Four out of five consecutive points fall in zone B or beyond, on the same side of the centreline.
- 4) Nine consecutive points fall on the same side of the centreline.

Whenever the data satisfies any of these conditions, the process is deemed out-of-control, and is flagged for further

examination to investigate the cause of variance. When the cause is identified, it can usually be removed such that the process is improved. It is worth noting at this point that the Quality Control Handbook states that the more rules that are incorporated, the higher the false-positive rate. Using these four rules, the false alarm rate is around 1 in every 91.75 observations [5]. The advantage however, is that using more rules gives a better indication of how controlled a process is.

Figure 2 shows clearly that this device is out-of-control, as several data points fall outside Zone A, satisfying Rule 1. However, the control chart obtained for an iPhone 3GS device (an updated version of the iPhone 3G), does not satisfy Rule 1 (as shown in Figure 3), meaning it is much more challenging to immediately ascertain the level of control for this process. The Western Electric Rules make it easier to infer the behaviour of the process. By reviewing the control chart against the four rules, it can be inferred that the device is in good statistical control, as none of the rules are satisfied. This means that the image acquisition process for this updated device has been vastly improved from its predecessor.

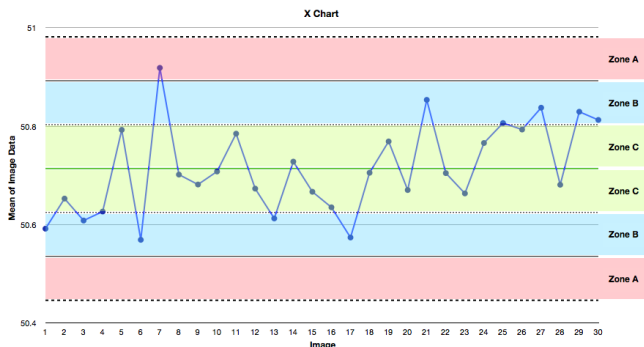


Fig. 3. X Chart for iPhone 3GS with highlighted zones.

The Western Electric Rules were applied to two iPhone 3G devices, one iPhone 3GS device, two Nokia N97 devices, and a Leica D-Lux 4 digital camera, as shown in Table II, where **X** denotes the existence of out-of-control observations (as they comply with the rule), and **✓** denotes statistical control. Both iPhone 3G devices are shown to comply with the same rules, whilst the updated iPhone 3GS yields a perfect level of control. This means that the image acquisition process has been improved, and thus a distinction can be made between images captured from either model. The Leica D-Lux 4 standalone camera also exhibits perfect statistical control.

TABLE II
COMPLIANCE TO WESTERN ELECTRIC RULES.

Device	Rule 1	Rule 2	Rule 3	Rule 4
iPhone 3G A	X	X	✓	X
iPhone 3G B	X	X	✓	X
iPhone 3GS	✓	✓	✓	✓
Nokia N97 A	X	X	✓	✓
Nokia N97 B	X	X	X	✓
Leica D-Lux 4	✓	✓	✓	✓

IV. EXPONENTIALLY WEIGHTED MOVING AVERAGE

In [4], X-Moving Range charts were used to demonstrate the effectiveness of SPC for illustrating the level of control observed from image data, for several digital cameras. These charts are typically associated with situations where individual samples are taken over time, and for this reason they are suited to the task. However, several authors have questioned the manner in which these charts are calculated as each observation is given an equal weighting regardless of when the observation occurred [12].

Exponentially Weighted Moving Average (EWMA) control charts offer an alternative approach to the same problem. Again, individual measurements are taken over time, but EWMA charts assign a weight to each data point such that each weight decreases exponentially from the present to the past. Thus, the EWMA chart more accurately reflects the recent behaviour of the process. The decrease of the weights can be calculated according to the following formula:

$$W_{t-i} = r(1-r)^i, \quad (1)$$

where r refers to a weighting factor that assumes values between 0 and 1, and W_{t-i} is the weight associated with data point X_{t-i} . If $r = 1$, then all the weight is given to the latest observation. In our implementation of EWMA control charts, r is set to 0.333 in order to yield a slower weighting decrement, without favouring the most recent observations too strongly.

In a similar vein to X-Moving Range control charts, a set of constants is required in order to construct the control charts, based around weighting factor r . A full list of constants for EWMA charts is listed in [11], and those associated with $r = 0.333$, are shown in Equation (2).

$$A^* = 1.342; D_1^* = 1.342; D_2^* = 1.780; d_2^* = 0.874 \quad (2)$$

where A^* is used for calculating the control limits for moving average charts, and D_1^* , D_2^* and d_2^* are used for constructing moving deviations charts. The moving average chart displays the accuracy of the process according to the process mean, and the moving deviations chart describes the process variability. Whenever a shift in the mean occurs, the exponentially weighted moving averages will gradually move towards the new process mean, whilst the standard deviations remain unaffected [11].

In order to calculate the control limits CL , UCL , and LCL for the moving averages chart A , the standard deviation S_X is first calculated for all observations.

$$S_X = \left[\sum_{t=1}^{30} \frac{(X_t - \bar{X})^2}{30-1} \right]^{\frac{1}{2}} \quad (3)$$

The control limits are then calculated using Equation (4).

$$CL_A = \bar{X}; UCL_A = \bar{X} + A^*S_X; LCL_A = \bar{X} - A^*S_X \quad (4)$$

The control limits for the moving deviations V_t charts are calculated from Equation (5).

V. CONCLUSION

$$CL_V = d_2^* S_X; UCL_V = D_2^* S_X; LCL_V = D_1^* S_X \quad (5)$$

Figure 4 shows the EWMA chart for an Apple iPhone 3G. The chart clearly shows the trend witnessed from the X-Moving range charts whereby a dramatic shift in variation occurs. However, as a result of assigning a weight to each observation, the EWMA chart is much more sensitive to this shift in the process mean. Subsequently, the control chart shows that the vast shift in variation is actually less abrupt than was inferred from X-Moving Range control charts. Figure 4 shows that the values increase slowly before reaching a certain point where the values start to slowly decrease. From a forensic standpoint this provides a little more useful information about the behaviour of the device.

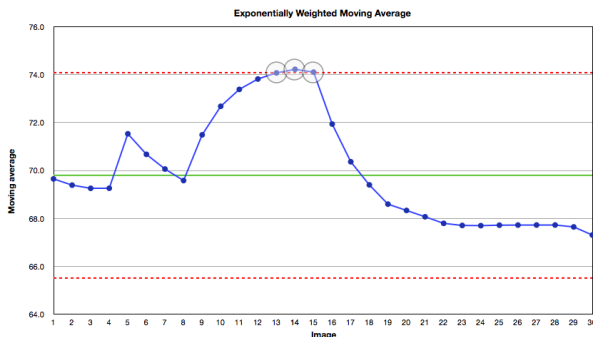


Fig. 4. EWMA Control Chart for iPhone A.

Figure 5 shows that the same trend can be observed from the EWMA chart obtained from a second iPhone 3G device. Again, the mean values reach a certain point, and then decrease steadily. The same pattern is witnessed for each iPhone 3G device that was tested, in addition to those reported in [4] which depicted a similar property). This similarity provides additional valuable information and confidence that this inherent compensating property exists in the iPhone 3G devices.

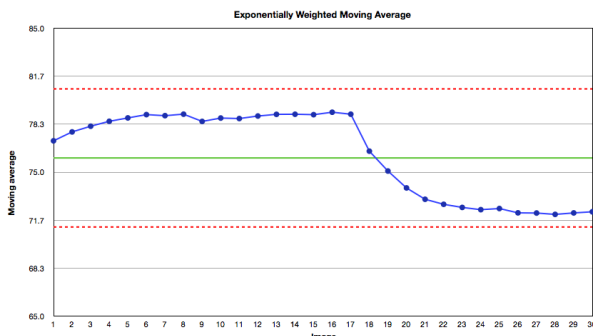


Fig. 5. EWMA Control Chart for iPhone B.

The potential of EWMA charts is promising for instances where, for example, a forensic expert might wish to measure the effect of increasing temperature on the device, over a period of time. By modifying r such that the observation weighting coincides with the increasing temperature, it should be possible to model the impact this has on the behaviour of the process.

In this paper, we have demonstrated the use of several SPC-based techniques for modelling and interpreting the individual behaviour of several digital cameras. After controlling the test environment, the resulting image data has been compared to that obtained from an uncontrolled environment, as reported in [4], where a similar behaviour was observed. When increasing the sample size to 30, the X-Moving Range control charts showed that the iPhone 3G devices were behaving similarly, aiding our confidence that this behaviour is a pure reflection of the make-up of the device, and not a response to changing environmental conditions.

The Western Electric Rules have also been applied to this application of SPC for identifying the level of control associated with each device. These rules highlighted that whilst the Apple iPhone 3G cameras are strongly out-of-control, the updated iPhone 3GS model was shown to be under perfect statistical control, meaning the cause of variation has been addressed for these devices. As iPhone 3GS devices do not inherit the same behaviour as their predecessor, a distinction can be made between image obtained from either model.

Finally, EWMA control charts have also been presented as an alternative approach to the X-Moving Range control charts discussed in [4] for any application where it is important to emphasise the most recent observations.

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] M. Kharrazi, H. T. Sencar, and N. Memon, "Blind Source Camera Identification," *International Conference on Image Processing*, vol. 1, pp. 709-712, 2004.
- [4] P. Bateman, A. T. S. Ho, and A. Woodward, "Camera Identification using Statistical Process Control Techniques for Anomaly Detection," *7th International Conference on Information, Communications, and Signal Processing*, to be presented at ICICS, Macau, 8-10 December 2009.
- [5] Western Electric Company, "Statistical Quality Control handbook," *Western Electric Company*, Indianapolis, Indiana, 1956.
- [6] Ford Motor Company, "Continuing Process Control and Process Capability Improvement," *Manual Published by Statistical Methods Office*, 1984.
- [7] 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.
- [8] 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.
- [9] A. L. Guillery, "Statistical Process Control in a Paper Mill," *Chemical Engineering Progress*, vol. 84, pp. 52-57, 1988.
- [10] W. A. Shewhart, "Economic Control of Quality of Manufactured Product," *American Society for Quality*, ISBN: 978-0873890762, 1931.
- [11] R. E. DeVor, T. Chang, and J. Sutherland, "Statistical Quality Design and Control: Contemporary Concepts and Methods," *Prentice Hall*, ISBN: 978-0023291807, 1992.
- [12] J. S. Oakland, "Statistical Process Control (Fifth Edition)," *Butterworth-Heinemann*, ISBN: 0-7506-3952-0, 2003.