Gottfried Wilhelm Leibniz Universität
Hannover, Germany

Studienarbeit

Statistical Monitoring of Image Data

Xilin Huang

Mentor: M. Sc.
Examiner: Prof. Dr.
Examiner: Prof. Dr.

June 29, 2021

The name of your protocol

Here write the background of your research.

Here write your research question.

Objectives
Here write your objectives in this master thesis.

Tasks and Timetable

- The first and the second months: literature review
- The second to the forth months: empirical study
- In the third month: mid-term reflection
- The forth and the fifth month: tuning the model(s)
- The sixth month: writing and presenting thesis
- The end: evaluation of performance (including thesis and presentation)

Tools
Algorithm:
Programming language: e.g., Python
Framework: e.g., Tensorflow
Datasets: e.g., Stanford drone datasets

PLEASE KEEP THIS PROTOCOL WITHIN ONE PAGE!

Statutory Declaration

I, YOUR NAME, declare that this master's thesis, and the work indicated herein have been composed by myself, and any sources have not been used other than those specified. All the consulted published or unpublished work of others have been clearly cited. I additionally declare that the work and master's thesis have not been submitted for any other previous degree examinations.

Xilin Huang Hannover, June 29, 2021

Eidesstattliche Erklärung

Ich, YOUR NAME, erkläre hiermit, die Arbeit selbstständig verfasst zu haben und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt zu haben. Alle Stellen der Arbeit, die wörtlich oder sinngemäß aus anderen Quellen übernommen wurden, habe ich als solche kenntlich gemacht. Die Arbeit wurde in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegt.

Xilin Huang Hannover, June 29, 2021

Acknowledgements

Here you can write your acknowledgements.

Abstract

Here write your abstract for your thesis. Please keep it accurate and clear within one page.

Contents

List of Acronyms

AI Artificial Intelligence

CNNs Conventional Neural Networks

DL Deep Learning

FNNs Feedforward Neural Networks

GPU Graphics Processing Unit

GRUs Gated Recurrent Units

IDE Integrated Development Environment

KLD Kullback-Leibler divergence

LIDAR Light Detection and Ranging

LSTM Long Short-Term Memory

ML Machine Learning

NLP Natural Language Processing

RNNs Recurrent Neural Networks

SGD Stochastic Gradient Descent

SDD Stanford Drone Dataset

List of Figures

List of Tables

1 Introduction

Here is the introduction. You need to write the following key points for your work. Keep each part concise and short. In total, this chapter should not be more than four pages.

The remainder of this paper is organized as follows. The real-time contrasts framework for image monitoring is introduced in Section 2. Section 3 evaluates the performance of the proposed method using simulations. Section 4 provides an experiment to apply the proposed method in an industrial environment. Finally, the conclusions and directions of future research are presented.

- Background
- Motivation
- Research gap
- Objectives
- Approach (in introduction chapter you do not necessarily need to write the results for your work)
- The structure of your thesis

2 Related Work

Here you need to write your related work and cite properly. For example, one of the most influential researchers in Deep Learning (DL) is Yann LeCun with the Nature Article *Deep learning* [**?**]. In total, you should refer to preferably approximately 50 papers (not less than 40 papers).

Here, it is highly recommended to start writing this part as soon as you start reading any papers. It will take you a lot of time to do so and can also help you track the papers that you have been reading.

3 Background

3.1 Discrete wavelet decomposition

3.2 Control chart

It is well known that statistical process control (SPC) is a powerful tool to prevent production of defects and nonconforming products. . Researchers believe that control charts are the most featured tool of SPC. Control charts can be classified into 2 categories of univariate and multivariate charts. In univariate control charts, . However, in multivariate control charts, one is interested in monitoring simultaneous changes in the parameter vector of an underlying multivariate distribution over time.

Statistical process control (SPC) is a powerful collection of problem-solving tools useful in achieving process stability and improving capability through the reduction of variability. SPC is one of the greatest technological developments of the twentieth century because it is based on sound underlying principles, is easy to use, has significant impact, and can be applied to any process [Montgomery, 2020]. Its seven major tools are

1. Histogram or stem-and-leaf plot
2. Check sheet
3. Pareto chart
4. Cause-and-effect diagram
5. Defect concentration diagram
6. Scatter diagram
7. Control chart

Of this seven tools, Control chart is probably the easiest yet effective tool to analyze the process stability. To best understand the concepts of Control chart, the classification of Control chart need to be clarified. Depending on the number of process characteristics to be monitored, there are two basic types of control charts. The first, referred to as a univariate control chart, is a graphical chart of one quality characteristic, in which one is interested in monitoring changes in the parameter of an underlying univariate distribution over time. The second is a multivariate control chart, which is a graphical chart of a statistic that fuses more than one quality characteristic to monitor simultaneous changes in the parameter vector of an underlying multivariate distribution over time.

More specificlly, the control chart is a graphical display, which plots the value of the quality characteristic that has been measured or calculated from a sample versus the sample number or versus time. Normally there are three lines in a control chart: a center line that correspond to the mean value for the in-control process. Two other horizontal lines, called the upper control limit (UCL) and the lower control limit (LCL). These control limits are chosen so that almost all of the data points will fall within these limits as long as the process remains in-control. A typical control chart is shown in Fig. 3.1.

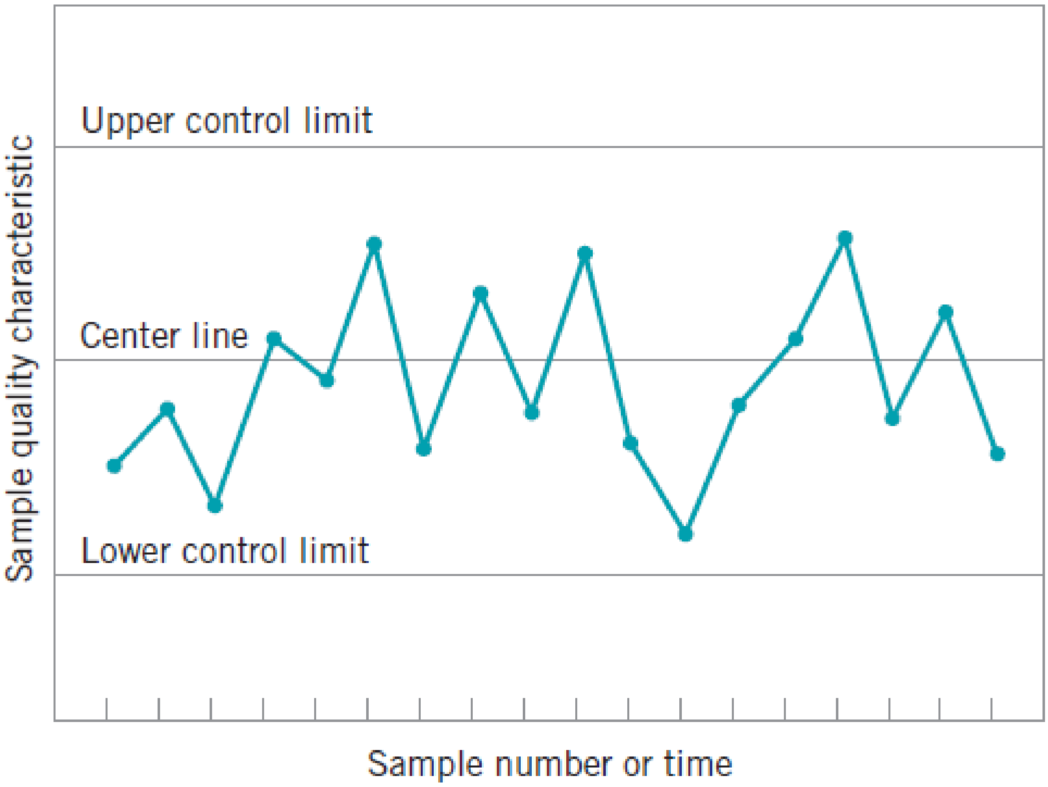

Figure 3.1: a center line that correspond to the mean value for the in-control process. Two other horizontal lines, called the upper control limit (UCL) and the lower control limit (LCL). Figure is adapted from [Montgomery, 2020]

3.2.1 Shewhart $\bar{X}$ and s control chart

In statistical quality control, the $\bar{X}$ and s chart is a type of control chart used to monitor variables data when samples are collected at regular intervals from a business or industrial process [Heckert et al., 2002].

In general, $\bar{X}$ and s control chart has following advantages:

1. The sample size (n) is relatively large ($n > 10$).
2. The sample size is variable.

The chart actually consists of two individual control charts: One ($\bar{X}$ control chart) to monitor the process mean and the other (s control chart) to monitor the process standard deviation. Firstly, we

There are two distinct phases of the control chart [Bersimis et al., 2007].

- Phase I: charts are used for retrospectively testing whether the process was in control when the first subgroups were being drawn. In this phase, the charts are used as aids to the practitioner, in bringing a process into a state where it is statistically in control.
- Phase II: control charts are used for testing whether the process remains in control when future subgroups are drawn. In this phase, the charts are used as aids to the practitioner in monitoring the process for any change from an in-control state.

In short, phase I that deals with estimating process parameters to ensure process stability using historical data, and phase II pertains to signal any out-of-control condition or shifts in the process parameters.

4 Methodology

The proposed framework combines feature extract methods and statistical process control (SPC) monitoring techniques to gradually infer a high-level statistic value in the cover of mobile phone image from the low-level representation of the scratches and stains and monitor the image base on statistic value. Firstly, the surface texture properties such as scratch and stain are decomposed into so-called statistical characteristics by mean of sliding-window method 4.1 and the wavelet transform 4.2. Then statistical approach, i.e., Shewhart control chart and Hotelling T^2 control chart, are utilized respectively to monitor mean statistic value and the mean statistic vector of a univariate and multivariate process, which can be used to judge the existence of scratch defects in the sample image.

4.1 Maximum variance of sliding-windows

We assume that the pixels on the surface of the phone's cover are homogeneous. Therefore, an idea of using the variance of pixel values to extract features representing the state of the surface is proposed.

The control charts are used spatially by moving a mask (or window) across the image and then calculating and plotting a statistic each time the mask is moved. The size of the mask depends on the expected size of the defects to be detected, with smaller defective regions requiring smaller mask sizes [Megahed et al., 2011].Inspired by this view, we move a ten by ten window (The size of the window is obtained by hand-tuning) across the image and calculate the variance of the pixel value each time the window is moved. The value with the largest variance among all windows in this image is taken as the desired statistic describing this image.

Since our goal is to employ the control chart to monitor the image data, we used S standard sample images as Phase I data to retrieve the mean value (m) and variance (σ) of the samples' statistic (maximum variance). In phase II, we will monitor whether the maximum variance of the incoming sample is within $m \pm \sigma$. If it is within this range, this sample will be considered qualified. Otherwise, it is unqualified.

4.2 Discrete Wavelet transform decomposition

The continuous wavelet transform is computed by changing the scale of the analysis window, shifting the window in time, multiplying by the signal, and integrating overall times. While in discrete wavelet transform (DWT) case, filters of different cutoff frequencies are used to analyze the signal at different scales. The signal is passed through a series of high pass filters to analyze the high frequencies, and it is passed through a series of low pass filters to analyze the low frequencies [Polikar et al., 1996].

The DWT [Fig. 4.1] analyzes the signal at different frequency bands with different resolutions by decomposing the signal into a coarse approximation and detail information, which is associated with low-pass and high-pass filters, respectively. In our case, we use Haar discrete wavelet transform as the basic function to perform signal decomposition so that an original image is decomposed into four coefficients: one low-pass filtering coefficients (approximation coefficients) and three high-pass filtering coefficients (detail coefficients, containing the horizontal (h), vertical (v), and diagonal (d) detail coefficients) at each level. The number of coefficients for approximation coefficients and detail coefficients is halved each time the level increase.

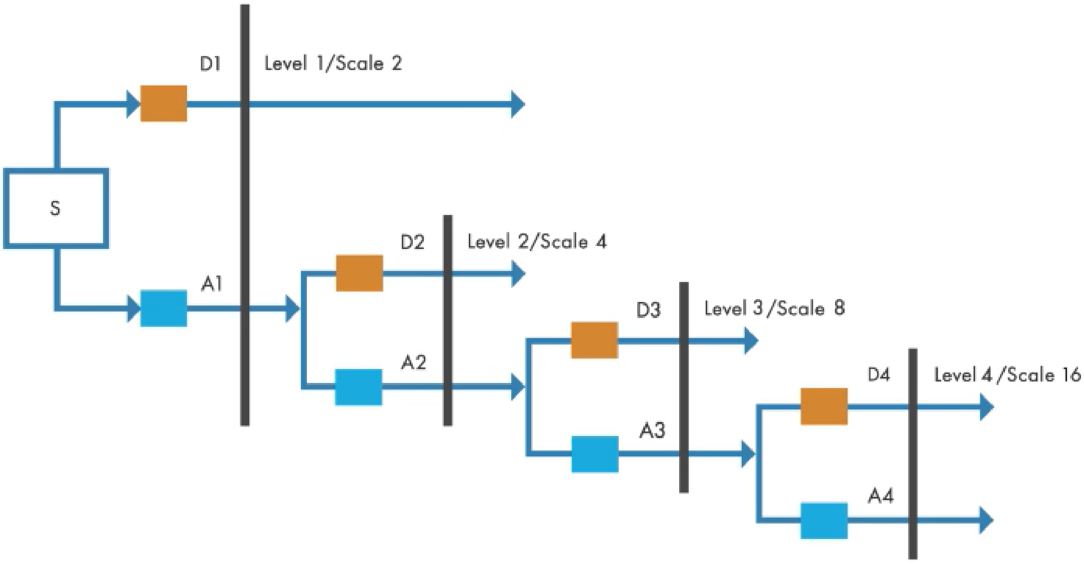

Figure 4.1: A general structure of DWT, the orange cube represent high-pass filter, the blue cube represent low-pass filter. Figure is adapted from MATLAB Tech Talks.

Since the analyzed images are in the form of 2-D, we need to perform the 2-D Haar wavelet transform by applying 1-D wavelet transform first on rows and then on columns. There is a built-in function in MATLAB Haart2, which performs a 2-D Haar wavelet transform. The lowest level (L_l) that can be obtained by the Haart2 transformation is

$$L_l = \log_2(min(row'sdimension, column'sdimension)) \tag{4.1}$$

If the row or column dimension of data is even, but not a power of two, the lowest level (L_l) that can be obtained by the Haart2 transformation is

$$L_l = \lfloor \log_2(\frac{min(row'sdimension, column'sdimension)}{2}) \rfloor \tag{4.2}$$

4.3 Wavelet decomposition based Hotteling T^2 control chart

In order to monitor the surface quality of the mobile phone's cover, feature statistics are needed to characterize the quality of the cover. The DWT is used to retrieved quality characteristics from image data since it decomposes the image data into some detailed coefficients, which contain horizontal, vertical, and diagonal coefficients. These processed coefficients can be fed into the control chart, in which the procedure can be judged whether it is under control by monitoring the value of the coefficients.

A RGB image has three frames: Red (R), Green (G), and Blue (B) frame. We apply Haar wavelet transform on each frame simultaneously. The coefficients of the final level (L_l) have L times filtered by the high pass filter, which means the amplitude of coefficients contains the information of the high-frequency signal in the original image. The higher the coefficients, the more likely the signal will be abrupt. The abrupt in the signal then correspond to the defects in the monitored products.

After we decompose an image of ($M \times N$) pixels, we get horizontal (h), vertical (v), and diagonal (d) coefficient matrix ($S \times S$) for each frame at final level (L_l), each coefficient matrix have S^2 coefficients,an image sample can have $3 \times S^2$ coefficients. **??** present tables indicating the recommended number of quality characteristics p = 2, 3, 4, 5, 10, and 20. It turns out that the coefficients of the diagonal coefficient matrix can best reflect the surface defects of various shapes (see experiment 6). Thus we first absolute all coefficients, which turn negative coefficients into positive coefficients without changing the value itself. After that, we take the maximal coefficient among the diagonal matrix and consider it the desired statistical characteristic of the corresponding frame. Further, three coefficients retrieved from three frames compose the multiple wavelet characteristic vector **x**. Finally, The Hotelling T^2 statistic

$$T^2 = (\mathbf{x} - \overline{\mathbf{x}})'\mathbf{S}^{-1}(\mathbf{x} - \overline{\mathbf{x}}) \tag{4.3}$$

in which $\overline{\mathbf{x}}$ and $\mathbf{S}$ be the sample mean vector and covariance matrix, respectively, of these observations, integrates the multiple wavelet characteristics $\mathbf{x}$ into a statistic value T^2 for each sample image. If this statistic value is larger than Upper Control Limit(UCL)

$$\mathrm{UCL} = \frac{p(m+1)(m-1)}{m^2 - mp} F_{\alpha,p,m-p} \tag{4.4}$$

in which m is the sample size, p is the number of quality characteristics, α is confident level, we are $(1-\alpha)$ confident that this sample is out of control. Vice versa, we are $(1-\alpha)$ confident that this sample is in control.

The output of the phase I (the sample mean vector $\overline{\mathbf{x}}$ and covariance matrix $\mathbf{S}$ of standard images as well as UCL) is used as the input of phase II. The images in phase II are first decomposed by Haar wavelet transform into 3×1 characteristic vector $\mathbf{x}$. Then by using

Equation 4.3 we can calculate the Hotelling T^2 statistic of this sample and compare it with UCL to judge if the sample is in control.

5 Dataset

Here you need to describe the dataset(s) that you use for your experiments. If you use some open–source data, please cite them properly.

Here is an example of the Stanford drone datasets (SDD) [**?**]. SDD provides a bird's-eye-view in various intersections on Stanford University campus.

frame Nr.	user ID	x	y	user type
834	0	819.5	29.5	2
835	0	819.5	29.5	2

Table 5.1: An example of the real data for SDD [**?**]. The first, second, third, fourth, and fifth column present the frame ID, user ID, x coordinates, y coordinates, and user type, respectively. The x and y coordinates are in pixels.

6 Experiments

Please wire the experiments you have been doing for your research problems in detail and in an understandable way. This is the part that will largely reflect the work you have been doing for your thesis and will also mainly decide whether you can pass the examination or not.

Please keep in mind that you backup your code properly and regularly to avoid any risks by accident. It is recommended to use some open–source platform such as GitHub[1], which is also very convenient for code sharing.

[1] `https://github.com`

7 Results and Discussion

Please write down the findings for the experiments that you have been doing. It is also better to describe the results in figures or tables than only plain text. But again, you also need to have adequate text to explain the underlying meaning for the results.

8 Summary and Outlook

8.1 Summary

Here you need to wrap up the thesis in very concise and short paragraph(s). This is normally a very frequent part that your readers take a first look.

You can also change this sub-chapter to conclusion if you can draw a conclusion based on the results you have found. However, please pay attention to the difference between conclusion and summary.

8.2 Outlook

Here you can point out some potential aspects or interesting directions for future work and improvement.

Bibliography

[Bersimis et al., 2007] Bersimis, S., Psarakis, S., and Panaretos, J. (2007). Multivariate statistical process control charts: an overview. *Quality and Reliability engineering international*, 23(5):517–543.

[Heckert et al., 2002] Heckert, N. A., Filliben, J. J., Croarkin, C. M., Hembree, B., Guthrie, W. F., Tobias, P., and Prinz, J. (2002). Handbook 151: Nist/sematech e-handbook of statistical methods.

[Megahed et al., 2011] Megahed, F. M., Woodall, W. H., and Camelio, J. A. (2011). A review and perspective on control charting with image data. *Journal of Quality Technology*, 43(2):83–98.

[Montgomery, 2020] Montgomery, D. C. (2020). *Introduction to statistical quality control*. John Wiley & Sons.

[Polikar et al., 1996] Polikar, R. et al. (1996). The wavelet tutorial.

Bibliography

[Bersimis et al., 2007] Bersimis, S., Psarakis, S., and Panaretos, J. (2007). Multivariate statistical process control charts: an overview. *Quality and Reliability engineering international*, 23(5):517–543.

[Heckert et al., 2002] Heckert, N. A., Filliben, J. J., Croarkin, C. M., Hembree, B., Guthrie, W. F., Tobias, P., and Prinz, J. (2002). Handbook 151: Nist/sematech e-handbook of statistical methods.

[Megahed et al., 2011] Megahed, F. M., Woodall, W. H., and Camelio, J. A. (2011). A review and perspective on control charting with image data. *Journal of Quality Technology*, 43(2):83–98.

[Montgomery, 2020] Montgomery, D. C. (2020). *Introduction to statistical quality control*. John Wiley & Sons.

[Polikar et al., 1996] Polikar, R. et al. (1996). The wavelet tutorial.