IE 423 Quality Engineering
Project Part 3, due January 9th, 2024

1. Quality Control on Images

You are a group of Quality Engineers in a linen manufacturer. "Linen is a textile made from the fibers of the flax plant. Linen is laborious to manufacture, but the fiber is very absorbent and garments made of linen are valued for their exceptional coolness and freshness in hot weather" (Source: Wikipedia).

"Automation of the visual inspection for quality control in production of materials with textures (tiles, textile, leather, etc.) is not widely implemented. A sophisticated system for image acquisition, as well as a fast and efficient procedure for texture analysis is needed for this purpose" (Source: Rimac-Drlje, Snježana, Drago Žagar, and Slavko Rupčić. "Adaptive Image Processing Technique for Quality Control in Ceramic Tile Production." Strojarstvo 52.2 (2010): 205-215.). Suppose you already have the image acquisition system and you are able to obtain pictures of linen (with a resolution of 512x512) as in Figure 1. This part of the project is about finding defects in the linen images.

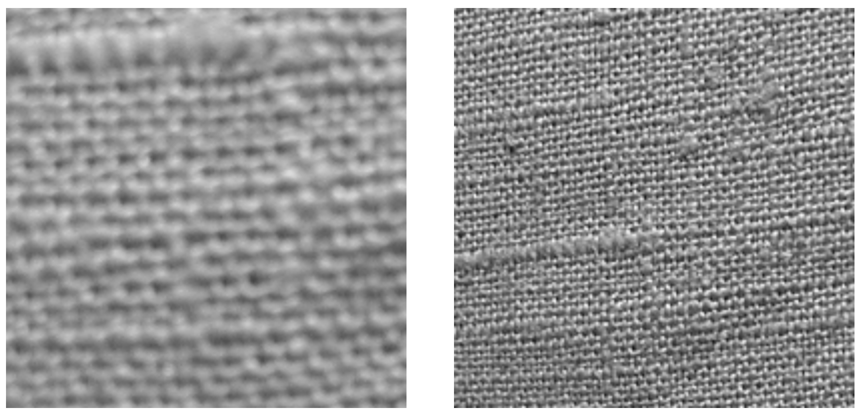

Figure 1: Sample linen images

2. Background - Images

Here is a background information about how a grayscale image is represented digitally on our computers. A grayscale image is basically a matrix where each matrix entry shows the intensity (brightness) level. In other words, when you take a picture with a digital camera, the image is represented by a numerical matrix where the matrix size is defined by the resolution setting of your camera. If your resolution setting is 1280x720, then your image is represented by 1280x720= 921600 pixel values (Actually that is why higher resolution provides better quality pictures).

When you have a color image, the image stores the information of multiple channels depending on the image type. The most famous one is RGB type where R, G and B stand for "red", "green" and "blue" respectively. Hence, you have a matrix as in greyscale images representing the intensity for each channel. Combining these matrices generates the color image.

You will be working on the images uploaded to Moodle for this task. There are several images in the provided zip file but you are expected to work on the "*.jpg" file named as "your_group_id x 5.jpg". In other words, if your group id is 5, you are expected to work on 5x5 = 25 → 0025.jpg file. Hence, each group will have their own image.

Before working on the tasks, please transform your color image to a greyscale one using an image editor. The other option is to use some package/module to perform the transformation. You will be working on greyscale images for this part of the project.

3. Some Strategies for Defect Detection

- A Baseline Defect Detection Approach from a Statistical Data Analysis Perspective

Potential steps for such an analysis are:

1. Suppose we are interested in the pixel value distribution of our image. Draw the histogram of the pixel values. Provide an appropriate probability distribution that fits well to the shape you observe (i.e. is it like Normal distribution, uniform or other?).
2. Assume that pixel values are following the distribution you have proposed in the previous task. Estimate the parameters using the data. (i.e. mean and variance for Normal distribution or min and max for uniform distribution and etc.).
3. Let's say the pixel values follow the distribution you proposed and its parameters are equal to what you have estimated in part 2. Identify the pixels that are out of the 0.001 probability limits. In other words, find a lower and upper bound that leave 0.001 of the observations on the smaller and larger side of the distribution respectively. Pixels that are out of these bounds should be identified. After finding those pixels, change the value of these pixels to zero (i.e. black color). Display the new image and original image in a plot. What are your observations? Comment on your findings.
4. Suppose we would like to perform the same operation on the patches of images (i.e. windows of certain size). When local structures are important, performing image operations on the patches might be important. Assume that your window size is 51x51 and you repeated what you have done in the first three tasks. Note that you do not need to draw each patch as requested in the previous task (i.e. task 3) but you are expected to mark the pixels for every patch. After finding those pixels, change the value of these pixels to zero (i.e. black color). Display the new image and original image in a plot. What are your observations? Comment on your findings.

- A Simple Defect Detection Approach from a Control Chart Perspective

In the previous proposal, you have performed quality control operations based on the pixel value distribution of the whole image and within the patches of the image. This proposal requires you to find out out-of-control pixel values considering the each row and each column of the matrix. Implicitly, we are imposing an assumption that the distribution of the pixels should be the same horizontally and vertically.

1. For each row of your image, construct an appropriate chart for monitoring the mean and variance. Find the pixels that are out of control. After finding those pixels, change the value of these pixels to zero (i.e. black color). Display the new image and original image in a plot. What are your observations? Comment on your findings.
2. Perform the same operations on the columns and comment on your findings.

- Your Proposal

In regular control charts, you observe a sequence of measurements (i.e. time series which is 1D). However, images are from a class of spatial data where you observe the data in 2D. You are expected to devise a control procedure to identify the regions of the image that might be problematic using statistical process control. In other words, you need to come up with a statistic that you think it can identify the irregularities in each image. As in all control charts, you are required to formulate an upper and lower control limit depending on the distribution of the statistic you monitor. As you may remember from the lectures, sometimes we tend to use multiple control charts and rules to identify the out of control points. In a complex monitoring like this one, use of multiple charts can help.

Regular control charts built as in the simple strategy assume that pixel values are independent of each other (i.e. there is a problem of autocorrelation). This is problematic especially for the type of image you are working on. When we work on an image which has a texture, the pixel values are somehow related (due to the definition of a texture). Given this background information, you are expected to devise a new strategy to detect defects.

4. Overall Comparison

Please evaluate "your proposal" on alternative images. For the selection of alternative images, generate 100 random integers between 2 and 196 (both included), take the first 5 available images based on the generated sequence. We have 146 images in total (not 196-2+1= 195 images), so select the first 5 available images for comparison.

5. Report & Code Documentation:

Your report should have the following format (Note that this report should not have any code integrated in the report, instead you can provide link/reference to the respective part of your code that will be provided as Appendix):

1. *Introduction:* What is linen? Why is it important to monitor processing of linens? What are the motivations regarding the use of images and identification of defects in linen manufacturing
2. *Background information:* What has been done in the literature regarding the process monitoring on linen?
3. *Approach:* Explain your approaches to this problem.
4. *Results:* Provide your results, comparison and discussion.
5. *Conclusions and Future Work:* Summarize your findings and comments regarding your approach. What are possible extensions to have a better approach?

Note that html pages are just to describe how you approach to the problem in the homework. You are also required to provide your R/Python codes separately in the repository so that anybody can run it with minimal change in the code. This can be presented as the script file itself or your notebook file (the one with *.ipynb file extension).

The last and the most important thing to mention is that academic integrity is expected! Do not share your code (except the one in your progress journals). You are always free to discuss about tasks but your work must be implemented by yourself.

Please mention if you utilize Large Language Models (i.e. chatGPT, BART and etc). You are expected to provide your prompt and resulting response of these models at the end of your reports as Appendix if you have utilized any of these tools.

As a fundamental principle for any educational institution, academic integrity is highly valued and seriously regarded at Boğaziçi University.

Best of luck, and remember – the journey of analyzing is as essential as the conclusions drawn!