9.3.2010

Lappeenranta University of Technology
Faculty of Technology Management
Department of Information Technology

The subject for the thesis was approved 21.04.2009

Demonstration Platform for Visual Quality Metrics

Bachelor's Thesis

Author:
Lauri Laaksonen

Supervisors:
Lasse Lensu
Tuomas Eerola

Abstract

Lappeenranta University of Technology

Faculty of Technology Management
Department of Information Technology

Lauri Laaksonen

Demonstration Platform for Visual Quality Metrics
Batchelor's thesis

2009

27 pages, 14 figures, 1 table

Examiners: Lasse Lensu
Tuomas Eerola

Keywords: DigiQ, Fusion of Digital and Visual Print Quality, print quality, image registration, image quality.

This thesis presents two graphical user interfaces for the project DigiQ - Fusion of Digital and Visual Print Quality, a project for computationally modeling the subjective human experience of print quality by measuring the image with certain metrics. After presenting the user interfaces, methods for reducing the computation time of several of the metrics and the image registration process required to compute the metrics, and details of their performance are given.

The weighted sample method for the image registration process was able to signifigantly decrease the calculation times while resulting in some error. The random sampling method for the metrics greatly reduced calculation time while maintaining excellent accuracy, but worked with only two of the metrics.

Tiivistelmä

Lappeenrannan Teknillinen Yliopisto

Teknillistaloudellinen tiedekunta
Tietotekniikan osasto

Lauri Laaksonen

Demonstration Platform for Visual Quality Metrics
Kandidaatintyö

2009

27 sivua, 14 kuvaa, 1 taulukko

Tarkastajat: Lasse Lensu
Tuomas Eerola

Hakusanat: DigiQ, Fusion of Digital and Visual Print Quality, tulostuksen laatu, kuvan rekisteröinti, kuvanlaatu.

Keywords: DigiQ, Fusion of Digital and Visual Print Quality, print quality, image registration, image quality.

Tässä tutkielmassa esitellään kaksi DigiQ - Fusion of Digital and Visual Print Quality – projektiin tehtyä graafista käyttöliittymää. Projektin tarkoituksena on mallintaa laskennallisesti ihmisen subjektiivista kokemusta tulostuksen laadusta mittaamalla tulostetta tietyillä mittareilla. Käyttöliittymien jälkeen annetaan menetelmät joidenkin esiteltyjen mittareiden ja mittareiden laskemiseksi tarvittavan kuvien rekisteröintimenetelmän laskenta-ajan vähentämiseen, sekä tutkitaan menetelmien toimivuutta käytännössä.

Kohtalaista virhettä aiheuttavalla kuvan rekisteröinnin nopeuttamiseksi kehitetyllä menetelmällä kyettiin vähentämään laskenta-aikaa huomattavasti. Mittareiden laskemiseen tarkoitettu random sampling – menetelmä ei aiheuttanut paljoa virhettä ja vähensi laskenta-aikaa huomattavasti, mutta toimi vain kahdella mittarilla.

Table of contents

Symbols

FR – Full reference
IFC – Information fidelity criterion
NQM – Noise quality measure
NR – No reference
RANSAC – Random sample consensus
SIFT – Scale invariant feature transform
SSIM – Structured similarity index
UQI – Universal quality measure
VIF – Visual information fidelity

1. Introduction

The thesis is organised as follows. Chapter 1 deals with the motivations for studying the print quality of the types of paper and the backgroud of the DigiQ project. Chapter 2 discusses the details of the demonstration platform and the approach to improving it's performance with weighted sampling and random sampling. Chapters 3 and 4 explain the test setting and results of the weighted and random sampling, and the following chapter 5 discusses the methods and their results.

1.1 Background

The modern printed media features an increasing amount of high resolution images and visualizations. The ability to print high quality images requires a high quality printing system, but also demands certain properties from the paper. Due to the effect of the paper's attributes on print quality, a way to computationally determine how humans experience the quality of the print on a certain paper becomes important when determining what kind of printing the paper is suitable for.

There are a number of metrics that can be used to measure certain aspects of an image that are related to print quality, but there are no guarantees that the results of these metrics correspond to human experience of the print quality.

1.2 DigiQ

The project DigiQ - Fusion of Digital and Visual Print Quality studies how the metrics for calculating image properties relating to print quality. As explained in [1], the data on the subjective human evaluations of print quality was gathered in a study by the University of Helsinki.

The subjective evaluation consisted of a number of students who were asked to study several sets of papers (16 or 21 papers, depending on the set) with an image printed on them. The students were then asked to rate the worst paper 1, the best paper 5 and the remaining papers 2,3 or 4 according to their perceived quality. The results of the metrics for calculating the properties relating to the print quality were compared to the data from the subjective evaluation to determine which metrics best correlated with the way humans perceive quality.

In [1], five metrics were found to correspond well with the subjective human experience of print quality: Structured similarity index (SSIM) [5], Noise quality measure (NQM) [6], Universal quality index (UQI) [4], Visual information fidelity (VIF) [7] and Information fidelity criterion (IFC) [7].

All the mentioned metrics compare a scanned (or distorted) image to the reference image that was printed and scanned (or transformed with some distorting transformation). As they require the reference image (as opposed to the technical and no-reference (NR) [9][10][11][12] metrics that, at most, require only a scan of certain test fields), they are referred to as full-reference (FR) metrics.

1.3 Objectives

The objectives of this project were to desing user interfaces for both computing the FR-metrics and the image registration process required for the computing of the metrics, and an interface for computing the NR-metrics.

The computation of the FR-metrics and especially the image registration process are very computationally heavy operations. The process of computing the FR-metrics for a single image can take dozens of minutes. Another objective for this project was to attempt to design methods for reducing the computation time of both operations while maintaining good accuracy.

The technical and NR-metrics are also computationally heavy operations, but since they are very heterogenious and measure very different properties of the print, each metric would most likely require a specific method for reducing it's computation time. Because of this, methods for improving the technical and NR-metrics were not addressed in this thesis.

2. Solutions

This chapter explains what steps were taken in the attempt to accomplish the objectives given in chapter 1.

2.1 User interface for image registration and calculating the FR and the technical metrics

As a part of this thesis, a demonstration setup for the image registration and the FR and technical metrics was created. The demonstration setup consists of a user interface for the computation of the technical and NR-metrics (**Figure 1**), and user interfaces for registering the images and computing the FR-metrics from the result.

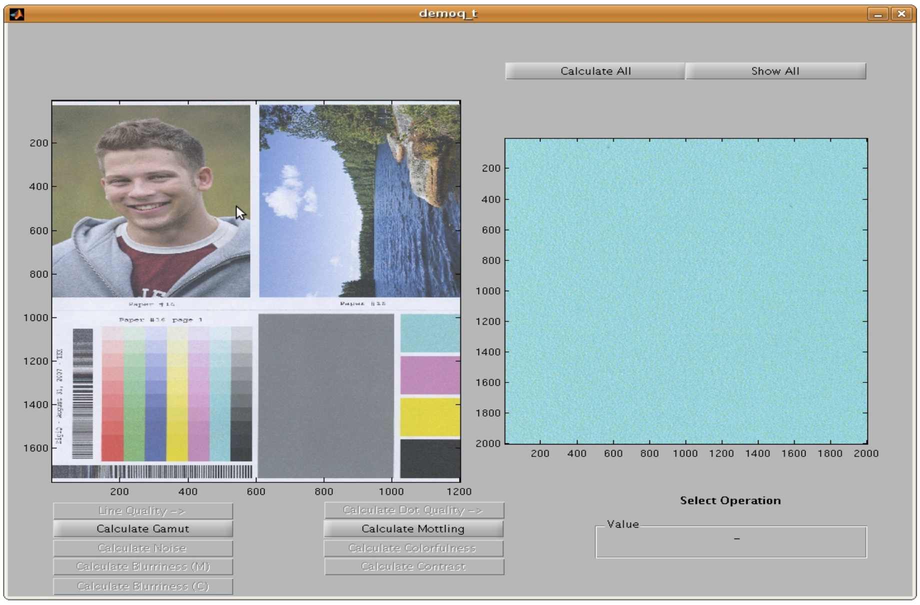

Figure 1: The user interface for the technical metrics

A previously created graphical user interface for image registration (see **Figure 2**) by Tuomas Eerola was modified to start the user interface for the calculation of the FR-metrics (see **Figure 3**) after the image registration process is complete.

Figure 2: User interface for image registration.

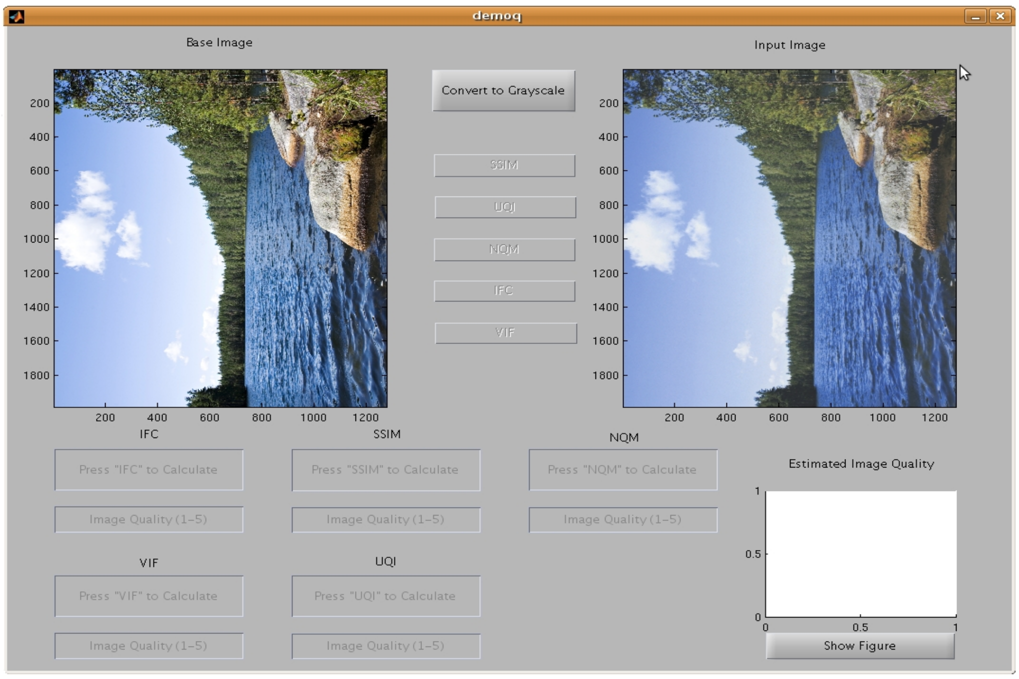

Figure 3: User interface for calculating the FR-metrics.

The process of computing FR-metrics from and image begins by starting the image registration user interface. The user then selects the scanned image (see "input image" in **Figure 2**) and it's reference image (see "base image" in **Figure 2**). The image resolutions are inserted and both images are filtered with a gaussian low-pass filter (see “descreening and scaling” in **Figure 2**, see [1] for explanation) with the same frequency.

After the filtering, the user pressess the “Register”-button and selects the region of interest (see “select ROI” in **Figure 2**) after the registration is complete. The “Crop”-button leaves only the regions of interest and “Continue” starts the user interface for computing the FR-metrics with the registered and cropped images.

Before the FR-metrics can be calculated, the images need to be converted to grayscale (see “Grayscale” in **Figure 3**). After the conversion, the user simply selects desired the FR-metric and the result is computed.

The technical and NR-metrics are computed by first clicking the desired field from the selection image (see the left image in **Figure 1**). The buttons for the metrics that can be computed from that field become active.

2.2 Improving the image registration process

The calculation time of the registration process can be reduced by using only a part of the image to be registered. As explained in [1], the image registration works by finding matching features with SIFT [3] on both the reference image and the scanned image and running RANSAC [2] on the features found.

While a greater number of features results in a more accurate image transform (the image transform to make the rotation and scale of the scanned image to match the ones of the reference image), not all possible features are needed for a successful registration.

There are, however, certain issues concerning the image registration process that cannot be ignored. While the greatest number of features can, in most images, be found in center of the image, the lack of features near the edges of the image causes errors around the edges of the registered image [1]. It can also be safely assumed that the image background contains very few features, if any.

When these two issues are considered, it is clear that some kind of algorithm must be programmed for selecting good samples from the scanned image, since selecting completely at random will most likely provide very unstable results.

These issues in mind, a method for reducing the calculation time of the registration process was designed. First, it runs the scanned image with Sobel edge detection filter to find the areas where most of the edges in the image (and therefore most of the image features) are. Since the scanned images are large due to high resolution requirements (the images must be scanned with sub-pixel accuracy to prevent scanning distortions affecting the registration process [1]), running the whole scanned image with the Sobel filter takes significant time.

Fortunately, filtering the whole scanned image is not required. Since we are only interested in the approximate regions that contain the most information, the scanned image can be scaled down for the edge detection. The regions containing the most information are determined by calculating the pixels left after the Sobel filtering in each region.

To counteract the issue of registration errors around the edges, the regions near the corners are given more weight than the ones near the center. As the Sobel filtering results in a binary image, weighting the edges can be achieved by simply leaving the image intensity values near the corners of the image as one and linearly decreasing the intensity values towards the center of the image.

When the regions with the most information from the weighted Sobel filtered image are determined, a number of the best regions are selected as samples and inserted on a blank image of the same size as the original scanned image. The registration process is then applied to the result and the reference image. The original scanned image is then transformed using the transformation matrix returned by the registration process.

2.3 Improving the process of calculating the FR metrics

Most of the FR algorithms calculate their metrics over the image in regions smaller than the whole image [5][4][7]. It would seem that running the algorithms over a number of random regions taken from the image and calculating the average of the resulting values would be a good method for reducing the calculation time. It is simply a matter of how many and how large regions are needed to reach an acceptable accuracy.

Unfortunately, the testing of the metrics with the random sampling method revealed that the inner mechanisms of most of the metrics (IFC, VIF and NQM) set certain restrictions to the random sampling method. These restrictions are explained in chapters **4.3.4**, **4.3.5** and **4.3.6**.

3. Testing the performance of the improved methods

The performance of the improved image registration method and the random sampling method for the FR metrics were tested with different variables. The purpose was to determine the optimal number and size of the regions, and if the methods can reduce the computation times without resulting in unacceptable loss in accuracy.

As errors in the image registration directly affect the performance of the FR metrics (to accurately compute the metrics the scanned image must have the same scale and rotation as the reference image), the accuracy of the improved registration method was considered more important than the reduction in the computation time.

The errors in the random sampling method for the FR metrics were, with accuracy still being an important factor, considered less critical. As the final estimate of the print quality is not taken from any single FR-metric [1], and as the FR-metrics different properties of the print [4][5][6][7], the final estimate is not so sensitive to errors in individual FR metrics.

3.1 Testing the improved image registration method

The improved registration method described previously has several parameters that can be varied: the scale of the image to be filtered with Sobel, the size of each region and the number of regions used. The effect of each variable on the accuracy of the improved registration process and the reduction in the computation time were tested by running the improved registration process on several sets of test images with different values for a certain variable at a time. The processing times and the transformation matrices received from the runs were recorded and analysed.

The accuracy of the improved method was tested on two sets of three images, with certain deformations applied to each set. In the first set, the reference images were rotated 1.2° clockwise and the images in the second set were scaled up, resulting in images 7% larger than the original reference images.

To measure the performance of the improved registration method, the recorded processing times and transformation matrices were compared to the results of a test run that registered the same sets of images using the original image registration process.

3.2 Testing the random sampling method for the FR metrics

The effect of region size and the number of regions on the performance of the FR metrics was tested by calculating the FR metrics over a set of previously registered images using varying region sizes and numbers of regions, and comparing them to the values of the metrics calculated over complete images.

To test the effect of the region size and the number of regions, a test run for both variables was performed, with the other variable fixed to a certain value. The interest was to determine if increasing one variable would provide signifigantly better results than increasing the other.

In an ideal case, a less than moderate increase in the size of the region would not signifigantly improve the result. This would justify using a large number of small regions, resulting in a greatly reduced calculation time and a more varied set of regions than using a small number of large samples would.

The values and the computation times from these test runs were analyzed to determine how the number and the size of the regions affect the accuracy, computation time and variance of the metrics. The variance of the values gained from the random regions are especially interesting since, even if the mean error in the values is at an acceptable level, occasional high changes in the values mean that the results are unstable and cannot be trusted.

4. Test results

Since each registration takes several minutes, the testing of the registration process was not as extensive as the tests on the FR metrics. The results of testing the registration process provided interesting data, but it should be treated as less accurate than the results of testing the FR metrics.

4.1 The image registration test results

Comparing the tranformation matrices returned by the improved image registration process revealed that the was no significant improvement when the lowest number and the size of the regions used (100 regions of ~0.03 % of the 1278x1987 pixel image) was increased.

Increasing the number and the size of regions does not significantly effect the accuracy, but increases the computation time. It would seem that the registration method is robust enough to work with the lowest tested number and size of the regions in the test sets.

In the final test, the 1278x1987 pixel test image was registered with 100 regions of size 75x75. The resulting tranformation matrix was compared to the matrix returned by registering the whole image. As **Table 1** shows, using the improved registration method resulted in maximum error of 0.0005 in the rotation part. The part of the matrix that handles the scale of the image has maximum error of 0.3168, which translates to ~33% relative error.

Transformation matrix from the original registering				**Transformation matrix from the improved registering**		
0.9998	-0.0264	0		0.9998	-0.0259	0
0.0262	0.9996	0		0.0261	0.9995	0
-53.7402	0.9670	1		-53.7942	0.6502	1

Table 1: Transfomation matrices from the original and the improved registration method

The error from the registration translates to approximately 2.5 % relative error when the FR metrics are applied (SSIM 2.38 %, UQI 3.82 %, IFC 2.42 %, VIF 1.48 %). The results from testing the NQM metric were not included, as NQM seemed to have trouble with the test set (the relative error of NQM for this particular test set was ~106 %, which is most likely an error when the signifigantly lower errors of the other metrics are considered).

The reduction in computation time was approximately 30 %. The test results suggest that the reduction in computation time will improve signifigantly when the size of the image increases. This is due to the fact that with a 7 % increase in the image size resulted in approximately 45 % reduction in the relative computation time.

4.2 The FR metrics test results

A number of problems arose when testing the performance of the random sampling method for calculating the FR metrics. Random sampling worked well with SSIM and UQI , but the other metrics, despite calculating their specific metric using only a part of the image at a time, performed certain operations or calculated arithmetics that made the algorithm break or give incorrect results.

4.2.1 Universal Quality Index

The first test of the Universal quality index (UQI) metrics was meant to determine which of the two variables (region size and the number of regions used) was more important for the metric to reach the correct result (the value gained from computing the metric from the whole image).

Increasing the number of samples results in a continuous approach (see **Figure 4**) towards the correct value. The metric is less likely to be affected by local features and the variance of the values quickly decreases as the number of samples used increases.

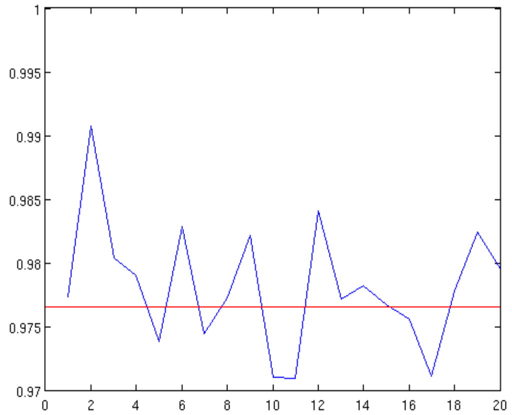

Figure 4: UQI with a fixed sample size and increasing the number of regions (correct value in red).

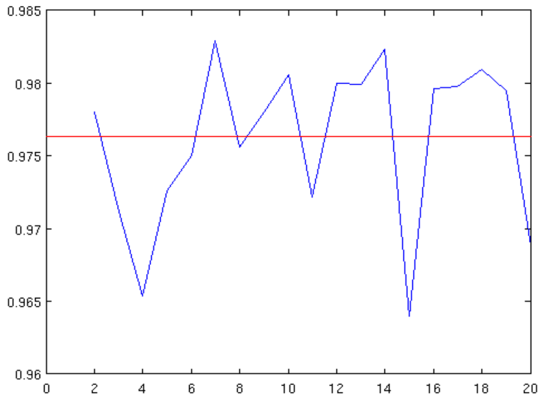

Figure 5: UQI with a fixed number of regions and increasing the region size (correct value in red).

Increasing the region size would make the metric approach the correct value in big steps (see **Figure 5**) instead of linearly progressing towards it. As the number of regions is relatively low, the metric is affected more by local features that have a significant effect on the value of the metric.

Using a large number of small regions provided more accurate result even though the percentage of the image used was approximately one fourth of the amount used with the sample size test run.

The second test was designed to provide information on the amount of regions that would provide a good reduction of computation time while generating only a minimal amount of error. As **Figure 6** shows, no significant decrease in the variation of the values returned by the metric can be achieved by increasing the number of samples after approximately 100 samples. This can be confirmed by studying **Figure 7**. After approximately 100 samples, there is no signifigant decrease of relative error when compared to the value from running the metric over the whole image.

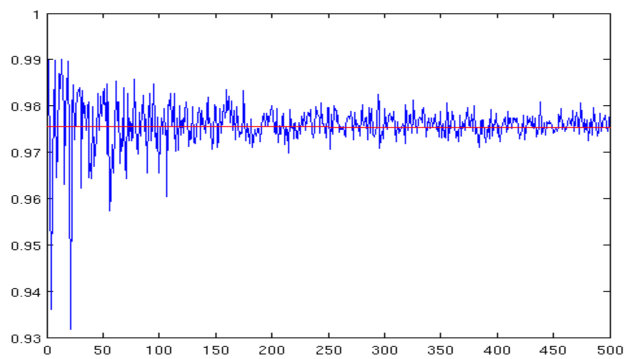

Figure 6: UQI with region size of 10x10 and increasing the number of regions (correct value in red).

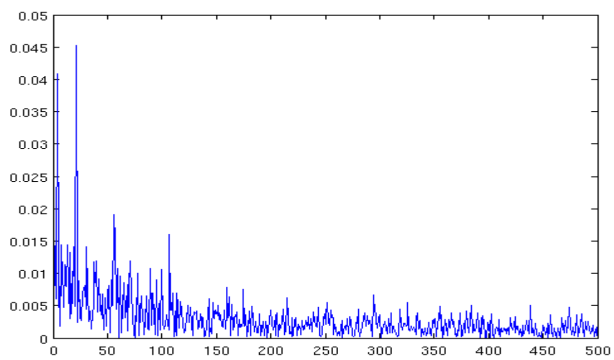

Figure 7: Relative error of UQI with sample size of 10x10 and increasing number of samples.

With sample size of 10x10 and using 100 samples, the average error was 0.5912 % with standard deviation of 0.0045. The average time required to compute UQI from one set of samples was 0.0715 seconds. Compared to the average of 12.8542 seconds it takes to calculate UQI from the whole image, random sampling reduces the calculation time by approximately 94 %. This suggests that random sampling works very well with UQI, as one can significantly reduce the computation time while maintaining excellent accuracy.

4.2.2 Structured Similarity Index

The testing process of Structured similarity index was identical to testing the UQI metric. As with UQI, the effect of sample size and the number of samples was tested to determine the better variable. If we compare **Figure 8** and **Figure 9**, it is clear that both increasing the sample size and increasing the number of samples produce results similar to the variables of UQI (see **Figure 5** and **Figure 4**).

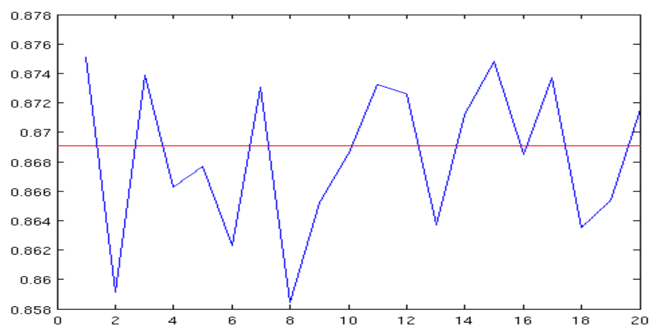

Figure 8: SSIM with a fixed number of regions and increasing the region size (correct value in red).

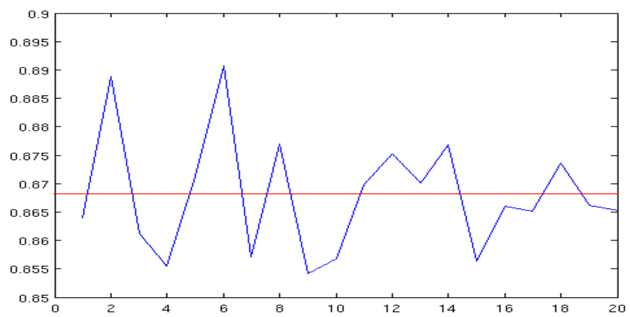

Figure 9: SSIM with a fixed region size and increasing the number of regions.

The effect of increasing the number of regions with sample size fixed to 10x10 can be seen in **Figure 10**. The values returned from computing SSIM over the regions behaved similarly to UQI (see **Figure 6**). The main difference was that the change in variation of the average values was significant up to 200 regions. Unsurprisingly, the change in the average error (**see Figure 11**) became insignificant after 200 samples.

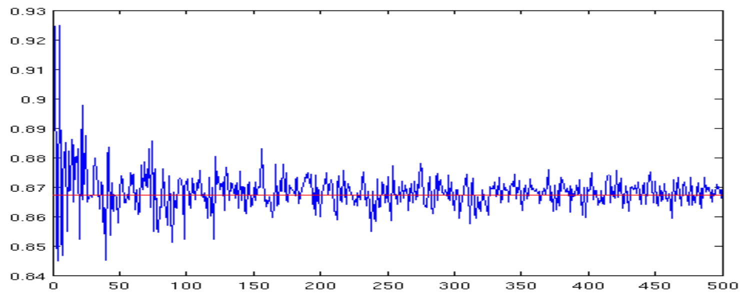

Figure 10: SSIM with a sample size of 10x10 and increasing the number of samples (correct value in red).

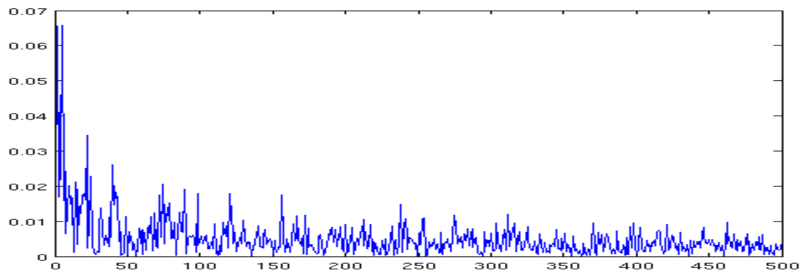

Figure 11: SSIM relative error with a sample size of 10x10 and increasing the number of samples.

The average error with 200 regions of size 10x10 was 0.722 % with standard deviation of the average values being 0.0047. The computation time dropped by approximately 97.7 %, as the average computation time over the whole test image was 9.6857 seconds and 0.2194 with random sampling. The test results show that random sampling results in a low error rate and an excellent reduction in computation time.

4.2.3 Information Fidelity Criterion

The gaussian pyramid algorithm the Information fidelity criterion utilizes [7] proved to be an issue with random sampling. The fact that the gaussian pyramid algorithm cannot be calculated from regions smaller than 80x80 pixels may limit the effective use of random sampling.

There is, however, an even bigger problem. The value that the metric computes is cumulative, resulting in increased return values as the size of the image or region increases. This increase in the value is not linear, so attempting to compute a relative value based on the image size would prove to be difficult. As IFC is only a part of this thesis, the attention that the issues presented would require is deemed excessive, and the issues will not be addressed in this thesis.

4.2.4 Visual Information Fidelity

As Visual information fidelity uses the same gaussian pyramid algorithm [7], it suffers from the same limitation to the region size. Considering this limitation and the fact that increasing the number of regions has proved to be the most effective method in both chapters **4.3.2** and **4.3.3**, the testing was limited to varying the number of regions with region size fixed at 80x80.

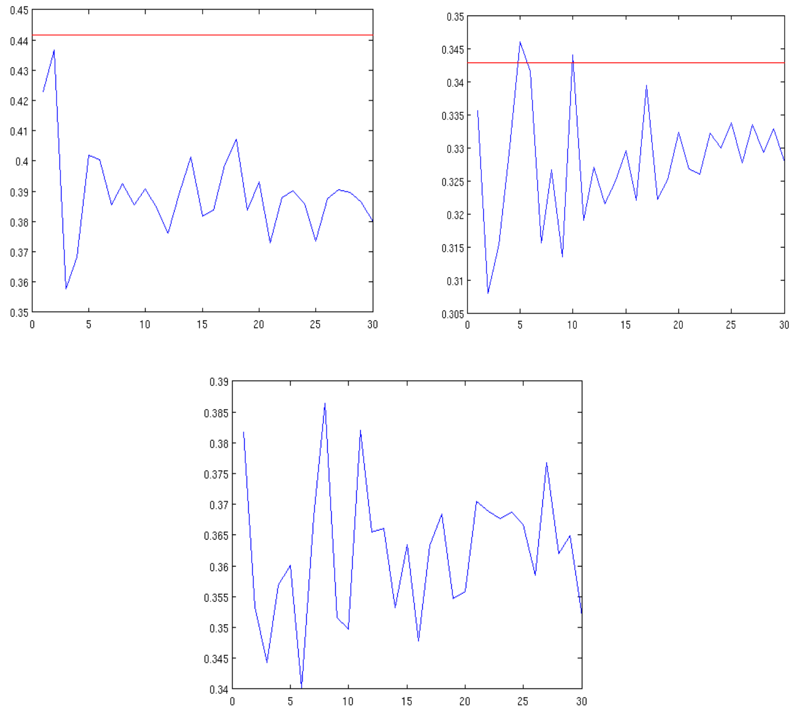

Figure 11: VIF from set 1, set 2 and set 3, with a region size of 80x80 and increasing the number of regions (correct value in red, correct value of the chart in set 3).

The performance of VIF with random sampling proved to signifigantly vary between the three test sets (see **Figure 11**). The changes in the variation of the average VIF values (see **Figure 12**) decrease as more regions are used, but the average error (see **Figure 13**) is still at unacceptable levels after 30 samples.

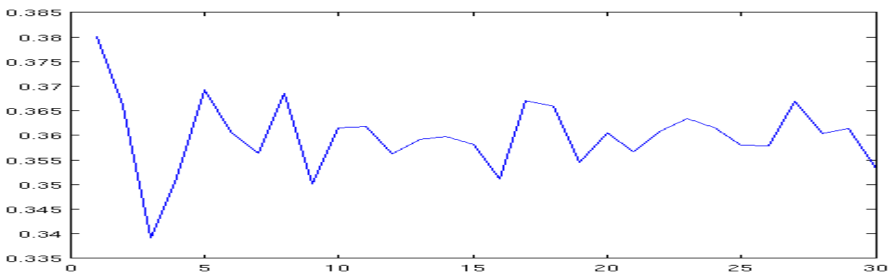

Figure 12: The average value of VIF over the three test sets (correct value off the chart).

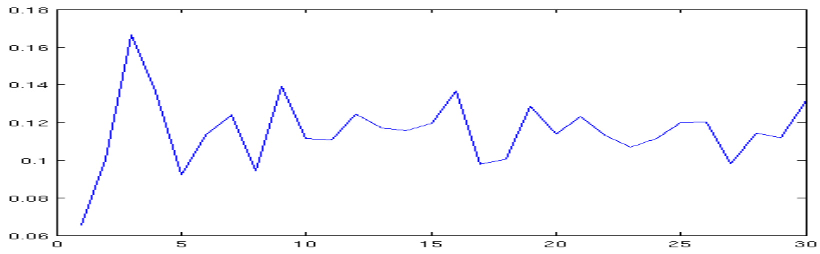

Figure 13: The average error of VIF over the three test sets.

This makes the use of random sampling with VIF problematic. The decrease in computation time gained from using random sampling becomes insignificant as the computation time over the whole image ranged from ~100 seconds to ~140 seconds depending on the image used (compare to **Figure 14**). While it is possible to use random sampling with VIF, it is not adviseable as any signifigant decrease in the computation time would result in unacceptable levels of error in the result.

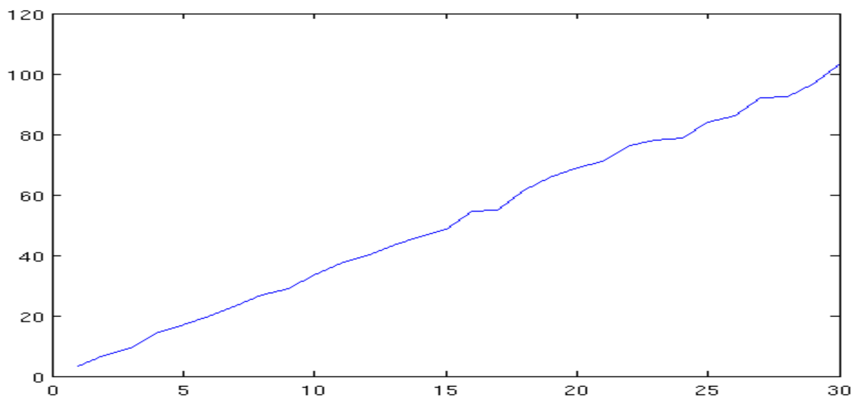

Figure 14: The effect of increasing the number of samples to the calculation time of VIF.

4.2.5 Noise Quality Measure

Testing the NQM metric revealed it to work very poorly with random sampling. The method that NQM uses to calculate the metric involves operations that tend to cause arithmetic errors when the size of the region is less than 200x200.

This means that each region is, at least, approximately 1.6 % of the whole image. The least amount of regions required for effective use of random sampling so far was 100 (see chapter **4.3.2**). With 200x200 regions this would amount to over one and a half times the whole 1278x1987 pixel test image and makes random sampling unlikely to provide meaningful results.

Furthermore, as with IFC, the return value of the metric depends on the size of the region. Even if this issue would be solved, the restrictions on the region size make the use of random sampling with NQM unwise.

5. Discussion

The weighted sampling method for the image registration process was able to register the input image approximately 30% faster than using the whole image for registering. Using the weighted sampling method causes approximately 2.5 % error in the FR metrics. This is acceptable if the required accuracy is relatively low, but should not be used with application with high accuracy requirements.

It is clear that if the performance of IFC, VIF and NQM are to be improved, a different approach needs to be taken. The tests proved that random sampling is simply not feasible for improving these metrics.

By testing SSIM and UQI metrics with random sampling, a good size and number of regions for both metrics were determined. By using a region size of 10x10 for both metrics, SSIM performed admirably with 200 regions and UQI with 100 regions. The computation time of both metrics could be reduced to only a few percents of the original times without losing much accuracy.

6. Conclusions

This document presented methods for improving the calculation times of image registration and calculation of FR metrics, user interfaces designed for registering images and computing FR-metrics, and a user interface for computing technical and NR-metrics.

It was shown that the method proposed for improving the image registration was able to considerably decrease the calculation time and maintaining reasonable accuracy.

The results from improving the calculation time of the FR metrics by using random sampling proved that only SSIM and UQI responded well to the method.

References

[1] - T. Eerola, J-K. Kämäräinen, L. Lensu, H. Kälviäinen, Framework for Applying Full Reference Digital Image Quality Measures to Printed Images, In the 16th Scandinavian Conference on Image Analysis, SCIA 2009, Oslo, Norway, 2009, Accepted for publication

[2] – M. Fischler, R. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Communications of the ACM, volume 24, no. 6, pp. 381-395, Jun. 1981

[3] – D. Lowe, Object recognition from local scale-invariant features, The Proceedings of the Seventh IEEE International Conference on computer vision, vol. 2, pp. 1150-1157, Sep. 1999.

[4] - Z. Wang, A. C. Bovik, A Universal Quality Index, *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81-84, March 2002

[5] - Z. Wang, A. C. Bovik, H. R. Sheikh, E.P. Simoncelli, Image quality assessment: From error visibility to structural similarity, IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, Apr. 2004

[6] - N. Damera-Venkata, T.D. Kite, W.S. Geisler, Image quality assessment based on a degradation model, Image Processing, IEEE Transactions on, Volume: 9, no. 4, pp. 636-650, Apr. 2000

[7] - H.R. Sheikh, A.C. Bovik, Image information and visual quality, Image Processing, IEEE Transactions on, Volume: 15, no. 2, pp. 430-444, Feb. 2006

[8] - T. Eerola, J-K. Kämäräinen, T. Leisti, R. Halonen, L. Lensu, H. Kälviäinen, G. Nyman, P. Oittinen, On Relation Between Human Visual Quality Experience and Measurable Quantities, IEEE Transaction on Systems, Man, and Cybernectics, Part A

[9] - ISO/IEC, 13660:2001(e) standard. Information technology – office equipment – measurement of image quality attributes for hardcopy output – binary monochrome text and graphic images, ISO/IEC, 2001

[10] - J.-E. Levlin, L.Söderhjelm, Pulp and Paper Testing, ser. Papermaking Science and Technology, Fapet Oy, 1999

[11] - A. Sadovnikov, L.Lensu, H. Kälviäinen, Automated mottling assessment of colored printed areas, in Proc. Of the 15th Scandinavian Conference on Image Analysis, Aalborg, Denmark, 2007, pp. 621-630

[12] - P. Oittinen, H. Saarelma, Kuvatekninen Laatu, Otatieto, 1992