A NO-REFERENCE IMAGE ENHANCEMENT QUALITY METRIC
AND FUSION TECHNIQUE

Thesis

Submitted to

The School of Engineering of the

UNIVERSITY OF DAYTON

In Partial Fulfillment of the Requirements for

The Degree of

Master of Science in Electrical Engineering

By

Jonathan Michael Headlee

UNIVERSITY OF DAYTON

Dayton, Ohio

May, 2015

A NO-REFERENCE IMAGE ENHANCEMENT QUALITY METRIC AND FUSION
TECHNIQUE

Name: Headlee, Jonathan Michael

APPROVED BY:

Eric Balster, Ph.D.
Advisor Committee Chairman
Associate Professor, Department of
Electrical and Computer Engineering

Frank Scarpino, Ph.D.
Committee Member
Professor Emeritus, Department of
Electrical and Computer Engineering

Vijayan Asari, Ph.D.
Committee Member
Professor, Department of Electrical and
Computer Engineering

John G. Weber, Ph.D.
Associate Dean
School of Engineering

Eddy Rojas, Ph.D., M.A., P.E.
Dean, School of Engineering

© Copyright by

Jonathan Michael Headlee

All rights reserved

2015

ABSTRACT

A NO-REFERENCE IMAGE ENHANCEMENT QUALITY METRIC AND FUSION TECHNIQUE

Name: Headlee, Jonathan Michael
University of Dayton

Advisor: Dr. Eric Balster

Image quality has always been an important aspect of the image processing field. Subjective quality is useful since images are a visual medium, but objective quality measures are needed because they are unbiased and can be used as parts of larger processing systems. Many image quality metrics exist that attempt to give an objective score to an image based on its likeness to a reference. These metrics work well if the reference is known and the test image is assumed to be a distorted version of the reference. However, in areas such as image enhancement, the reference image is generally worse than the test image and measuring likeness between the two is not a good indication of visual quality. A no-reference image enhancement quality metric is proposed in this paper that uses three factors to score images: lightness, contrast, and noise. It has been shown in literature that certain ideal ranges for lightness and contrast exist, and image enhancement techniques tend to push an image towards these. The metric gives each pixel in an image a score based on its neighborhood statistics. An image fusion technique is also proposed that fuses multiple enhanced images into one based on the local scores obtained from the no-reference metric. It is shown that this fused image scores higher using the no-reference metric and also has superior visual quality.

ACKNOWLEDGMENTS

I would like to thank everyone who encouraged and helped me during my time at UD as a graduate student, especially the following people:

- **Dr. Eric Balster:** For advising me throughout my time at the University of Dayton and always being there to assist me.
- **Dr. Frank Scarpino and Dr. Kevin Priddy:** For providing me with imagery to study and test with throughout my research.
- **Kerry Hill, Bill Turri, and the entire UDRI ADDA Lab Team at the Air Force Research Laboratory at Wright-Patterson Air Force Base:** For giving me this opportunity to pursue a master's degree and helping me grow as a professional engineer.
- **Friends and family:** For the constant support throughout my education at UD.

TABLE OF CONTENTS

ABSTRACT	iii
ACKNOWLEDGMENTS	iv
LIST OF TABLES	vii
LIST OF FIGURES	viii
 I. INTRODUCTION	 1
1.1 Existing Image Quality Metrics	2
1.2 Image Enhancement Methods	4
1.3 Organization	7
 II. IMAGE QUALITY	 10
2.1 Statistics of Images	11
2.2 Impact of Noise on Image Statistics	14
 III. NO-REFERENCE IMAGE ENHANCEMENT METRIC	 19
3.1 Mean Quality Factor	20
3.2 Variance Quality Factor	21
3.3 Noise Quality Factor	22
3.4 Overall Quality Factor	23
 IV. NO-REFERENCE METRIC RESULTS	 25
 V. IMAGE ENHANCEMENT FUSION PROCESS	 30
5.1 Image Fusion Motivation and Overview	30
5.2 Description of Proposed Weighting System	31

VI. IMAGE FUSION TECHNIQUE RESULTS	33
VII. CONCLUSIONS AND FUTURE WORK	45
7.1 Conclusion	45
7.2 Future Work	45
BIBLIOGRAPHY	47

LIST OF TABLES

4.1	No-reference Image Metric Results	25
6.1	No-reference Metric Scores for Original, Enhanced, and Fused Images	33

LIST OF FIGURES

1.1	Original Image Before Enhancement	6
1.2	GHE Enhanced Image	7
1.3	CLAHE Enhanced Image	8
1.4	LTSN Enhanced Image	9
2.1	Optimal Visual Lightness and Contrast	11
2.2	Original False Color Image	13
2.3	GHE False Color Image	14
2.4	CLAHE False Color Image	15
2.5	LTSN False Color Image	16
2.6	Original Image + False Color	17
2.7	GHE Image + False Color	18
3.1	No-reference Image Enhancement Metric Flowchart	19
4.1	Boat Images. Top left: Original, $M_s = 0.6593$, Top right: GHE, $M_s = 0.7666$, Bottom left: CLAHE, $M_s = 0.7266$, Bottom right: LTSN, $M_s = 0.7701$	26
4.2	Door Images. Top left: Original, $M_s = 0.7896$, Top right: GHE, $M_s = 0.7801$, Bottom left: CLAHE, $M_s = 0.8276$, Bottom right: LTSN, $M_s = 0.8135$	28

4.3	Lakehouse Images. Top left: Original, $M_s = 0.5425$, Top right: GHE, $M_s = 0.6803$, Bottom left: CLAHE, $M_s = 0.6256$, Bottom right: LTSN, $M_s = 0.5955$	29
5.1	Image Fusion Flowchart	31
5.2	Weighting Algorithm Flowchart	32
6.1	Boat Images. Top left: Original, $M_s = 0.6593$, Top middle: GHE, $M_s = 0.7666$, Top right: CLAHE, $M_s = 0.7266$, Bottom left: LTSN, $M_s = 0.7701$, Bottom middle: Fused, $M_s = 0.8115$	34
6.2	Boat Images. Top left: Original, Top middle: GHE, Top right: CLAHE, Bottom left: LTSN, Bottom middle: Fused	35
6.3	Lakehouse Images. Top left: Original, $M_s = 0.7896$, Top middle: GHE, $M_s = 0.7801$, Top right: CLAHE, $M_s = 0.8276$, Bottom left: LTSN, $M_s = 0.8135$, Bottom middle: Fused, $M_s = 0.8401$	36
6.4	Lakehouse Images. Top left: Original, Top middle: GHE, Top right: CLAHE, Bottom left: LTSN, Bottom middle: Fused	37
6.5	Door Images. Top left: Original, $M_s = 0.5425$, Top middle: GHE, $M_s = 0.6803$, Top right: CLAHE, $M_s = 0.6256$, Bottom left: LTSN, $M_s = 0.5955$, Bottom middle: Fused, $M_s = 0.7049$	38
6.6	Door Images. Top left: Original, Top middle: GHE, Top right: CLAHE, Bottom left: LTSN, Bottom middle: Fused	39
6.7	Cave Images. Top left: Original, $M_s = 0.8140$, Top middle: GHE, $M_s = 0.7890$, Top right: CLAHE, $M_s = 0.8718$, Bottom left: LTSN, $M_s = 0.8375$, Bottom middle: Fused, $M_s = 0.8742$	40
6.8	Cave Images. Top left: Original, Top middle: GHE, Top right: CLAHE, Bottom left: LTSN, Bottom middle: Fused	40
6.9	Fireplace Images. Top left: Original, $M_s = 0.6722$, Top middle: GHE, $M_s = 0.7609$, Top right: CLAHE, $M_s = 0.7567$, Bottom left: LTSN, $M_s = 0.7622$, Bottom middle: Fused, $M_s = 0.8031$	41

6.10	Fireplace Images. Top left: Original, Top middle: GHE, Top right: CLAHE, Bottom left: LTSN, Bottom middle: Fused	42
6.11	Mountain House Images. Top left: Original, $M_s = 0.8553$, Top middle: GHE, $M_s = 0.8247$, Top right: CLAHE, $M_s = 0.8684$, Bottom left: LTSN, $M_s = 0.8266$, Bottom middle: Fused, $M_s = 0.8914$	43
6.12	Mountain House Images. Top left: Original, Top Middle: GHE, Top right: CLAHE, Bottom left: LTSN, Bottom middle: Fused	43
6.13	Island Images. Top left: Original, $M_s = 0.8397$, Top middle: GHE, $M_s = 0.7636$, Top right: CLAHE, $M_s = 0.8570$, Bottom left: LTSN, $M_s = 0.8202$, Bottom middle: Fused, $M_s = 0.8698$	44
6.14	Island Images. Top left: Original, Top Middle: GHE, Top right: CLAHE, Bottom left: LTSN, Bottom middle: Fused	44

CHAPTER I

INTRODUCTION

Image quality metrics have been used in the image processing community throughout its history. Images are used to convey information in a visual way, so a way to measure the quality is very important. Evaluating imagery can be done both subjectively and objectively. Subjective analysis is the human factor and is important, but varies on the individual and what they look for in a “good” image. For example, some may find any amount of noise to be damaging, while others can live with noise as long as it does not affect what information can be seen. Subjective analysis is a good tool to score images but it is inexact, not repeatable, and slow, which means it cannot be used in real time computer systems. For these reasons, objective measures need to be used to score and compare images. Many of these objective metrics fall under the full-reference (FR) category, where a degraded image is compared to a known ideal reference. This works well in some areas of image processing such as image compression, since the result of image compression is a distorted image worse than the original. However, in other fields, no reference image exists or it does not help in the calculation of image quality. One such field is image enhancement, where the goal is to enhance an image to look visually appealing to humans or increase accuracy in tracking or other artificial intelligence tasks. There is no ideal optimal image, since the starting image is already degraded. The enhanced image should not have distortions, so many of the traditional FR metrics do not make sense to use. Some no-reference (NR) metrics exist in literature but have limited uses.

Many of them require previous knowledge on the type of distortions that will be present in an image, which does not help for image enhancement. This paper introduces a localized no-reference image enhancement metric to score various image enhancement techniques. These scores are then used to fuse multiple enhanced images together to create an overall superior enhancement process. The following sections outline some basic information about currently existing image quality metrics and enhancement methods, and are followed by a section detailing the structure of this paper.

1.1 Existing Image Quality Metrics

The most common and widely used image quality metrics are full-reference. These metrics generally compare the images pixel by pixel to get a quality score that shows how much the degraded image deviates from the ideal. For some areas, such as image compression, FR metrics such as mean square error (MSE) and the related peak signal-to-noise ratio (PSNR) are used since the ideal image is readily available. These metrics are well known and can be found in [1]. MSE is given as

$$MSE = \frac{1}{N} \sum_{x=0}^{N-1} (\hat{I}(x) - I(x))^2, \quad (1.1)$$

where N is the total number of pixels in the image, $\hat{I}$ is the degraded image, and I is the ideal image. MSE gives the average squared “error” in an image, where error is how much a pixel in the degraded image differs from the same pixel in the ideal. It is apparent that a “good” degraded image would have a lower MSE than a “bad” one. PSNR is related to the MSE and is given in the decibel scale as

$$PSNR = 20 * \log_{10} \left(\frac{MAX_I}{\sqrt{MSE}} \right), \quad (1.2)$$

where MAX_I is the maximum pixel value for the image (255 for typical 8-bit imagery). PSNR relates the power of a signal to the noise corrupting it. For image compression, the artifacts introduced represent the noise. As stated before, the lower the MSE the better. Since MSE is on the denominator of the PSNR calculation, a high PSNR will correspond to a “good” degraded image.

MSE and PSNR are useful techniques to see how much a degraded image differs from an original ideal. They are conceptually simple and easy to implement so they are widely used. However, they have issues discussed in detail in [2]. The metrics do not take into account any of the intricacies of human visual perception, and because of that their scores for images might not match what a human would think.

Visual information fidelity (VIF) is a more recent full-reference image quality metric that uses natural scene statistics and a human visual system model to measure image quality [3]. The metric quantifies the amount of information contained in a reference image and measures how much of the information can be found in the degraded image. This new way of defining quality based on image information instead of signal fidelity produces results more in line with what a human would perceive.

The three metrics described above all use an ideal, undistorted reference image and compare this to the distorted, degraded image. However, a reference image is not always available. No-reference image quality metrics are a difficult task because there is nothing to compare to. Many no-reference metrics make assumptions about the type of distortion that will be in an image and assess the quality based on that distortion. For example, in [4], a no-reference algorithm is designed to account for JPEG compression distortions, namely blocking artifacts and blur. These types of no-reference metrics have a limited range of uses. Since they make assumptions about the underlying distortion in a test image, they will not perform well if other distortions are present. Furthermore, assuming a distortion is present also makes the assumption that the test image is “worse”. This is a problem for image enhancement since the test image should actually look better.

1.2 Image Enhancement Methods

The goal of image enhancement is to make an image “better”. Of course, “better” is a subjective term. An image could be enhanced to look better visually to humans, or it could be enhanced in such a way to improve performance in other image processing tasks. A common operation within image enhancement is contrast enhancement, or “deshadowing”. Global histogram equalization (GHE) is a well known image enhancement method that increases global contrast of an image [1]. It does this by adjusting intensities to be more spread out on an image histogram. In theory, a narrow range of intensity values should be mapped to a wide range in the output. Areas that are very bright or very dark should see an increase in visibility due to the contrast stretching. GHE is a popular technique due to its ease of implementation. However, its global nature can lead to overamplification of noise, an undesirable characteristic which hampers subjective visual quality as well as potential detection algorithms.

Contrast limited adaptive histogram equalization (CLAHE) is a specialized form of adaptive histogram equalization (AHE) [5]. AHE algorithms transform each pixel of the image using a transformation function calculated from a local neighborhood surrounding the pixel. This local approach to histogram equalization performs better than GHE on images with significant variations in intensity. However, basic AHE has a similar problem of enhancing noise in homogenous areas of the image. If the local neighborhood is homogenous, the algorithm will map this narrow range of intensity values to a wide range and even initially small amounts of noise will be apparent. CLAHE prevents this overamplification of noise by setting a clip limit on the contrast level. The resulting histogram is clipped so that any high peaks are evenly distributed among all histogram bins. This prevents overamplification of noise and results in visually appealing imagery.

The locally tuned sine nonlinear transform (LTSN) is an enhancement technique that was developed both for enhancement of video frames in surveillance systems and an enhancement technique

for images under non-uniform lighting conditions [6]. A nonlinear transfer function is used to enhance the intensity of the image based on a multi-level Gaussian function. Dark areas are made brighter while bright areas are reduced in intensity. Contrast enhancement is then done, followed by a color restoration (ignored for grayscale imagery). LTSN does well in images with both very dark and very bright areas. Many contrast enhancement algorithms (such as GHE) do well in dark areas but can overexpose lighter areas. LTSN avoids this with its nonlinear nature.

The three described enhancement techniques all have similar functions but perform differently, as can be seen in this example. Figure 1.1 shows an image that has not been enhanced and its histogram. The histogram shows how many pixels of each intensity are in the image. This is a standard 8 bit grayscale image, so there are 256 intensity levels. Most of the pixels have an intensity value below 50 which corresponds to how dark the image is. It is obvious that the image has underexposed areas, so enhancing it should increase what is visible. Figure 1.2 shows the same image after performing GHE.

As shown in Figure 1.2, an increase in contrast is observed due to GHE processing. More detail can be seen on the boat, but the sky and water have become overexposed. Figure 1.3 shows the image after CLAHE.

There is much more contrast in the water and sky compared to the GHE image, but some areas of the boat are darker. The CLAHE histogram shows that more pixels are at a higher intensity level than the original but they are not as spread out as in the GHE image. Figure 1.4 shows the image after performing LTSN.

The LTSN image gives good detail at the rear of the ship but contrast is lost between the sky and water. Each of the above image enhancement techniques have differing strengths and weaknesses. There is no one best algorithm for all situations, which means evaluating techniques should take

Figure 1.1: Original Image Before Enhancement

place on an image by image basis. This furthers the need of an automated objective metric to evaluate image enhancement techniques.

Figure 1.2: GHE Enhanced Image

1.3 Organization

The organization of this thesis is as follows. Chapter II discusses statistics of images and noise, the basis for the rest of the work. Chapter III discusses the no-reference image enhancement method.

Figure 1.3: CLAHE Enhanced Image

Chapter IV shows the results of the proposed enhancement method. Chapter V details the fusion enhancement process, where multiple enhanced images are scored and then combined based on the

Figure 1.4: LTSN Enhanced Image

scores. Chapter VI shows the results of this fusion process. Chapter VII concludes the paper and details future work.

CHAPTER II

IMAGE QUALITY

Analyzing an image's quality can be done subjectively and objectively. However, it is a challenging task to objectively evaluate image quality, especially when evaluating image enhancement algorithms. Full reference measures such as MSE and PSNR do not make sense when evaluating enhanced images. They measure pixel by pixel differences in a reference and test image. As seen in the above figures, enhancement techniques drastically change these pixel values. The images would have poor MSE and PSNR values, but this does not mean the enhancement techniques are bad. It simply means MSE and PSNR are poor indicators of quality for enhanced images. VIF does a better job with enhanced images because of its use of information as a basis for image quality. In a contrast enhanced image the perceptual quality should be higher, and the VIF's use of a HVS model signifies this. However, there is still a reliance on the original reference image, which may not always be available. Existing no reference metrics fare no better for measuring enhanced image quality. As stated before they must have prior knowledge on the type of distortions in the image. However, an enhanced image ideally should not have any distortions. Using a no reference metric designed to score images based on a certain type of distortion (such as JPEG artifacts) [4] is not helpful in evaluation of image enhancement.

2.1 Statistics of Images

In [7], Rahman, Jobson and Woodell discuss the statistics of visual representation. They say that a good image has a combination of high regional lightness and contrast. Lightness can be measured with the local mean of the image. It is simply a measure of how bright the image is in the specified region. Contrast is measured with the local standard deviation. This can be thought of as a measure of the content in the specified region. The local means and standard deviations are averaged to get an overall mean and standard deviation score. Note that this overall standard deviation is not the same as the standard deviation if it is calculated for the entire image. Rather, it is the mean of all of the local standard deviations. The authors hypothesize that good visual representations of images converge to specific statistical values, seen in Figure 2.1. These values are for typical

Figure 2.1: Optimal Visual Lightness and Contrast

8-bit imagery where the highest intensity value is 255. An increase in quality is seen from left to right in the visually optimal region. Intuitively, this plot makes sense. A visually optimal image

would not be too dark or too bright, but would lie somewhere in the middle. Additionally, having a high contrast will make objects in the image distinguishable, which is important for both humans and feature detectors that need to classify objects. Low contrast could result from a uniform area in the image. Enhancing a uniform area should not increase contrast by much, if any. This is not optimal in the sense that no real information can be obtained from this area. For example, if most of the image is of a clear sky, enhancing the image will not bring out any new details. This indicates that there are limits to how much an image can be enhanced, and that not all images can reach the optimal region for lightness and contrast. Low contrast could also result from an area that is too bright or dark. This seems to say that lightness and contrast have a relationship: enhancing an image should increase both. Increasing lightness can in effect “reveal” areas of an image where previously there was nothing. This revealed area should have an increased contrast now that objects are visible. With more objects visible, more information can be seen in the image. Figures 2.2-2.5 illustrate lightness and contrast on the boat image before and after enhancing it with GHE, CLAHE, and LTSN. The idea for these false-color images comes from [8] and were created to demonstrate how enhancement techniques improve both lightness and contrast. Individual pixels in the image are colored based on their regional statistics. A purple pixel in the false color image represents one whose neighborhood has poor lightness and contrast (a local mean and standard deviation below 100 and 35, respectively). A blue pixel has a neighborhood with poor contrast but an acceptable lightness. Red pixels are the opposite of blue, with poor lightness but acceptable contrast. Green pixels are the most optimal, with both acceptable lightness and contrast.

The original false color image shows that large regions of the image are poor. The image is originally underexposed in some areas, so the majority of the ship is lacking in both lightness and contrast. The sky is colored blue which represents it has a high mean but low contrast. This makes sense because although the sky is bright, there is little detail (there are clouds, however, they do not

Figure 2.2: Original False Color Image

provide enough contrast with the sky to get into the visually optimal region for contrast). Green areas are present where the water meets the sky and where the sky meets the boat. There is an obvious difference between these surfaces, and the sky is bright enough for to see the edge well. Enhancing the original image using GHE gives a much brighter output, which shows up in the false color image. More detail is able to be seen, especially at the back of the boat. However, contrast is lost between the sky and water, and also between the water and the railing of the boat. The CLAHE image in Figure 2.4 has much better detail in the water and railing, but is not as bright as the GHE image. This causes a loss in some lightness and contrast on the boat. The LTSN false color image is similar to CLAHE but has a few noticeable differences. More detail is visible in the background of the LTSN image at the cost of foreground contrast. For underexposed images like this example, increasing lightness also has the benefit of generally increasing contrast. All of the enhancement

Figure 2.3: GHE False Color Image

techniques improved on the original image in some way, and they all performed differently for the different aspects of the image.

In this paper, the local standard deviations are calculated for imagery and are then squared to get variances. Looking at the square of the standard deviation gives information in terms of the energy contained in the image. This idea is used in PSNR, described in Chapter I. It uses MSE to get a ratio between power and noise, and the squared error is used because loss is historically looked at in terms of power loss.

2.2 Impact of Noise on Image Statistics

The previous discussion and results work well when there is little noise in the imagery. However, as stated in Section 1.2, GHE has a tendency to overamplify even small amounts of noise. An

Figure 2.4: CLAHE False Color Image

example of this can be seen in Figures 2.6 and 2.7. The original image is extremely dark, with only the scoreboard and banners illuminated. There is a small amount of noise in the background, but its not visible. Using GHE on the image exposes the background well but also amplifies the noise. Evaluating this enhanced image is very difficult. Subjectively, one might say there is much more detail in the GHE image and the noise does not affect object recognition. Others may find the noise extremely offensive and rate this image lower than the original. Noise causes problems with objective quality measures as well. Large portions of the GHE image are either red or green, which means the contrast is high. However, there is no useful information in the sky, where only noise is enhanced. When noise is this pronounced in an image, it can skew image statistics. Fortunately, there are ways to estimate noise found in literature [9, 10]. The method described in these papers makes use of the wavelet transform to estimate noise in the larger process of de-noising imagery. In

Figure 2.5: LTSN False Color Image

this process, a multiscale wavelet transform is done to get multiple sets of wavelet coefficients. A sub-band of these coefficients are then used to give the noise estimation as

$$\sigma_{\eta} = \frac{\text{Median}(|\tilde{\lambda}_{hh,0}[\cdot]|)}{0.6745}, \quad (2.1)$$

where σ_{η} is the estimated standard deviation of the noise level and $\tilde{\lambda}_{hh,0}[\cdot]$ are the wavelet coefficients in the high-high band of the finest (0th) scale. The median of these coefficients are used because most of the useful information in this band is contained in a few large coefficients. Using the mean would skew the estimation, but the median ignores the useful coefficients and gets an accurate level of the noisy ones. An estimate of the noise is another useful statistic when discussing image quality, and as shown in the next chapter, used in the proposed image quality metric.

Original

Original False Color

Figure 2.6: Original Image + False Color

GHE

GHE False Color

Figure 2.7: GHE Image + False Color

CHAPTER III

NO-REFERENCE IMAGE ENHANCEMENT METRIC

With the knowledge that a good visual representation of an image converges to certain values of lightness and contrast, and that noise should be minimized, a no-reference image enhancement metric is proposed. This metric is given by

$$M(x) = \alpha Q_{\mu}(x)Q_{\eta}(x) + (1 - \alpha)Q_{\sigma}(x). \quad (3.1)$$

A visualization of the method is shown in Figure 3.1. The proposed image enhancement metric

Figure 3.1: No-reference Image Enhancement Metric Flowchart

utilizes three separate localized quality factors, $Q_\mu(x)$, $Q_\sigma(x)$, and $Q_\eta(x)$, representing lightness, contrast, and estimated noise content, respectively. α is a tuneable scaling parameter less than or equal to 1 and can be used to change the impact that mean or standard deviation have on the score. For the testing shown in this thesis, α is kept at 0.5. The quality factors are calculated on a per-pixel basis and combined together to form the score matrix. This matrix gives the metric score for each pixel. These scores are averaged to finally get to the overall image quality score. The three quality factors used in the metric will be analyzed in further detail in the following sections.

3.1 Mean Quality Factor

The mean quality factor of Equation 3.1 is given by

$$Q_\mu(x) = \frac{\mu_m(x)^2 - (\mu_i(x) - \mu(x))^2}{\mu_m(x)^2}, \quad Q_\mu(x) \in [0, 1], \quad (3.2)$$

where $\mu(x)$ is the local neighborhood mean for pixel at spatial location x , $\mu_i(x)$ is the ideal mean, and $\mu_m(x)$ is the maximum possible difference between $\mu_i(x)$ and $\mu(x)$. The local neighborhood mean is defined as

$$\mu(x) = \frac{1}{N} \sum_{k \in n(x, N)} I(k). \quad (3.3)$$

$I(k)$ is a pixel value in spatial location k . $n(x, N)$ represents the neighborhood of N pixels centered about spatial location x . Typically, N is set to 101 to encapsulate the local area. The ideal mean is defined as

$$\mu_i(x) = \begin{cases} \mu_h, & \mu(x) > \mu_h \\ \mu_l, & \mu(x) < \mu_l \\ \mu(x), & \text{else} \end{cases}, \quad (3.4)$$

where $[\mu_l, \mu_h]$ is the range of ideal image mean. This comes from [7] and Figure 2.1, but is tweaked and set to $[\mu_l, \mu_h] = [110, 155]$ after testing. Note that the ideal ranges for mean and standard deviation can be modified at any time, and the provided ranges are what were used to generate results shown in this paper. In Equation 3.2, $\mu(\cdot)$ is subtracted from $\mu_i(\cdot)$ to get the distance between the local and ideal means. This difference value is then subtracted from $\mu_m(\cdot)$, and the entire thing

is divided by $\mu_m(\cdot)$ to get Q_μ in the range $[0, 1]$. As $\mu(\cdot)$ strays farther from the ideal, the overall mean quality factor approaches 0. Conversely, if $\mu(\cdot) = \mu_i(\cdot)$, $Q_\mu(\cdot) = 1$. Having a range for the ideal mean allows an adaptive selection of an ideal for the current neighborhood statistics. For example, if a pixel has a very high local mean, it means that the image is bright in that area. If the mean is higher than μ_h , the ideal is set to μ_h . The opposite is true for dark areas with low local means. If $\mu(\cdot)$ is within the range defined by $[\mu_l, \mu_h]$, the ideal is set to $\mu(\cdot)$. Doing this ensures that the ideal somewhat is in line with the current neighborhood. If only one value is chosen for the ideal, say 100, a pixel with local mean 150 would be incorrectly scored low, even though 150 is still within the visually optimal region. The last thing to relate is $\mu_m(\cdot)$, which is given as

$$\mu_m(x) = \begin{cases} \mu_i(x), & \mu_i(x) \geq 2^{B-1} \\ 2^B - \mu_i(x) - 1, & \text{else} \end{cases}. \quad (3.5)$$

B is the bitdepth of the imagery and is set to 8. $\mu_m(\cdot)$ gives the maximum possible difference between $\mu_i(x)$ and $\mu(x)$.

3.2 Variance Quality Factor

The variance factor of Equation 3.1 is structured similarly to Equation 3.2 and is given as

$$Q_\sigma(x) = \frac{\sigma_m(x)^2 - |\sigma_i(x)^2 - \sigma(x)^2|}{\sigma_m(x)^2}, \quad Q_\sigma(x) \in [0, 1], \quad (3.6)$$

where $\sigma(x)$ is the local neighborhood standard deviation for pixel at spatial location x , $\sigma_i(x)$ is the ideal standard deviation, and $\sigma_m(x)$ is the maximum possible difference between $\sigma_i(x)$ and $\sigma(x)$. Note that these values are squared in the calculation. This gives variance terms, which better match human perception of visual quality and information loss. The local neighborhood standard deviation is defined as

$$\sigma^2(x) + \sigma_\eta^2(x) = \frac{1}{N} \sum_{k \in n(x, N)} (I(k) - \mu(k))^2. \quad (3.7)$$

$I(k)$ is a pixel value in spatial location k . $n(x, N)$ represents the neighborhood of N pixels centered about spatial location x , and $\mu(k)$ is the local mean. Note that since noise in the image is modeled

as additive, the calculation in Equation 3.7 is the sum of the image variance $\sigma^2(\cdot)$ and noise variance $\sigma_\eta^2(\cdot)$. Calculation of $\sigma_\eta^2(\cdot)$ is detailed in Section 3.3, and once found can be subtracted from the result of Equation 3.7 to get the image variance $\sigma^2(\cdot)$. The ideal standard deviation is given as

$$\sigma_i(x) = \begin{cases} \sigma_h, & \sigma(x) > \sigma_h \\ \sigma_l, & \sigma(x) < \sigma_l \\ \sigma(x) & \text{else} \end{cases}, \quad (3.8)$$

where $[\sigma_l, \sigma_h]$ is the range of ideal image standard deviation. Again, this is taken from [7] and Figure 2.1 and is set to $[\sigma_l, \sigma_h] = [50, 78]$ after testing. The variance quality factor acts just as the mean quality factor, with a score ranging from 0 to 1. $\sigma_m(x)$ is the maximum possible difference between $\sigma_i(x)$ and $\sigma(x)$ and is given as

$$\sigma_m(x) = \begin{cases} \sigma_i(x), & \sigma_i(x) > \frac{2^B-1}{2\sqrt{2}} \\ \sqrt{(2^{B-1}-1)^2 - \sigma_i(x)^2}, & \text{else} \end{cases}, \quad (3.9)$$

where B again is the bitdepth of the imagery and set to 8.

3.3 Noise Quality Factor

An estimate of noise in imagery is given in Equation 2.1. This technique for estimating image noise is modified to create a local noise estimation, given as

$$\sigma_\eta(x) = \frac{\text{Median}(|\tilde{\lambda}_{hh,0}[x]|)}{0.6745}, \quad (3.10)$$

where $\text{Median}(x)$ is the output of a median filter of neighborhood $n(x, N)$ applied to the first level of coefficients of the wavelet transform of the image in the high-high band, denoted as $\tilde{\lambda}_{hh,0}[x]$. This equation gives an estimation of the standard deviation of the noise for the current pixel's neighborhood. This estimate is used both in the previous section to indirectly calculate image variance and as a variable in the noise factor of Equation 3.1, given as

$$Q_\eta(x) = \left(1 - 3 \left(\frac{\sigma_\eta(x)}{\mu(x)}\right)^2\right), \quad Q_\eta(x) \in [0, 1]. \quad (3.11)$$

The fractional term is an estimate of the signal to noise ratio. The constants are there to ensure the overall noise factor falls within the range $[0, 1]$. It is assumed that in a worst-case scenario there

could be all noise and no signal. Furthermore, this noise would vary uniformly. Assuming the noise varies uniformly from 0 to η , the variance of the noise is

$$\sigma_{\eta}^2(x) = \frac{\eta^2}{12}, \quad (3.12)$$

and that the mean squared of the noise is

$$\mu_{\eta}^2(x) = \left(\frac{\eta}{2}\right)^2 = \frac{\eta^2}{4}. \quad (3.13)$$

These results come from the well known properties of uniform distributions. Plugging them into the signal to noise estimation, the result is

$$\frac{\sigma_{\eta}(x)^2}{\mu(x)^2} = \frac{\left(\frac{\eta^2}{12}\right)}{\left(\frac{\eta^2}{4}\right)} = \frac{1}{3}. \quad (3.14)$$

This result is used in Equation 3.11 to get $Q_{\eta}(\cdot)$ to fall within the correct range. If there is no noise present, σ_{η} will be 0, which makes the entire $Q_{\eta}(\cdot)$ term go to 1.

3.4 Overall Quality Factor

The overall quality factor for each pixel is given in Equation 3.1. Note that each piece of that equation falls in the range $[0, 1]$, where 1 is maximal quality. With the α parameter present, the highest total score is also 1. For a pixel to get an overall score of 1, it must have an ideal neighborhood mean and variance with no noise degradation. As the mean and standard deviation vary from ideal, and a large amount of additive noise is present, $M(\cdot)$ approaches 0. Note that the noise factor is multiplied by the mean factor but not the variance factor. This is because a decorrelation has already been done between the variance and noise, described in Section 3.2. When calculating the variance factor the noise variance is first subtracted, so in a way the noise is already taken into account.

To obtain an overall quality score for an image, $M(\cdot)$ is merely averaged over all pixels.

$$M_s = \frac{1}{N} \sum_x M(x), \quad (3.15)$$

where N is the total number of pixels in the image, and M_s is the overall quality score.

CHAPTER IV

NO-REFERENCE METRIC RESULTS

To test the proposed no-reference metric, various images of differing original quality and lightness conditions were used. These images were enhanced with GHE, CLAHE, and LTSN, and scores were generated for both the original and enhanced versions. Figure 4.1 shows the original and enhanced “Boat” images, Figure 4.2 shows the original and enhanced versions of the “Door” image, and Figure 4.3 shows the same for the “Lakehouse” image. Table 4.1 gives the results in a tabulated form. The boat image has been discussed previously in Chapter II. It can be seen that each en-

Table 4.1: No-reference Image Metric Results

	Boat	Lakehouse	Door
Original	0.6593	0.7896	0.5425
GHE	0.7666	0.7801	0.6803
CLAHE	0.7266	0.8276	0.6256
LTSN	0.7701	0.8135	0.5955

hancement technique improves upon the original image and consequently has a higher score. LTSN has the highest score, so it can be said that it does the best job enhancing both lightness and contrast with minimal noise. The lakehouse is interesting because the original actually scores higher than the GHE enhanced image. There is a bit more lightness on the actual house in the GHE image, but some contrast is lost in the trees. Otherwise the original and GHE images look very similar, and the

Figure 4.1: Boat Images. Top left: Original, $M_s = 0.6593$, Top right: GHE, $M_s = 0.7666$, Bottom left: CLAHE, $M_s = 0.7266$, Bottom right: LTSN, $M_s = 0.7701$

closeness of the scores indicates this. CLAHE has the highest score for the lakehouse with good lightness of the house and contrast in the water. The door image has GHE as the highest scoring enhancement technique. It is the brightest and detail can be seen; however, detail is lost in the windows. This shows a problem with using the mean pixel score to get a global score. Other techniques would have higher scores in the windows but globally GHE outclasses them, so it gets the overall highest score. The door image also has some noise apparent in the enhanced versions. LTSN does very well in enhancing the overexposed windows, but has the highest amount of noise, especially in the ceiling. This detracts from its overall score and is why it is rated lowest of the enhancement techniques. Again, this noise is only apparent in certain areas of the image. LTSN would have the highest pixel scores in areas including the windows.

These results show the usefulness of the no-reference metric to compare existing image enhancement techniques. However, it has repeatedly been discussed and shown that different enhancement techniques do well in different areas. The next chapter details a method to fuse the techniques to create an overall superior image. More in-depth results of the metric are included after this discussion of the fusion method.

Figure 4.2: Door Images. Top left: Original, $M_s = 0.7896$, Top right: GHE, $M_s = 0.7801$, Bottom left: CLAHE, $M_s = 0.8276$, Bottom right: LTSN, $M_s = 0.8135$

Figure 4.3: Lakehouse Images. Top left: Original, $M_s = 0.5425$, Top right: GHE, $M_s = 0.6803$, Bottom left: CLAHE, $M_s = 0.6256$, Bottom right: LTSN, $M_s = 0.5955$

CHAPTER V

IMAGE ENHANCEMENT FUSION PROCESS

5.1 Image Fusion Motivation and Overview

Previously it has been shown that different image enhancement techniques perform differently depending on the image or even regions of the same image. Recall that the proposed no-reference image enhancement metric gives an overall quality score based on the average of each pixel's quality score. Looking at the overall score is sufficient for talking about the metric, but there are uses for the individual pixel scores as well. Looking at these scores one can see where a particular metric does well and where it does not. It is likely that different enhancement techniques will do better and worse in different areas. This leads to the idea of fusing techniques together to maximize quality, with Figure 5.1 showing the process. The process starts with the original, un-enhanced image. From there, N different enhancement techniques are done to generate N enhanced images. Note that N is an arbitrary number; in this paper three enhancement metrics have been discussed but any number could be added or removed. Also, note that the original image can be put into the weighting system. This could be useful if every enhancement technique actually makes the image worse: in this case the original should have high weights. The N enhanced images are then put into a weighting system, which gives each a pixel by pixel weight for the output. The weights are used to fuse the images and create a composite that maximizes the good of each technique and minimizes the bad.

Figure 5.1: Image Fusion Flowchart

5.2 Description of Proposed Weighting System

Assume N different image enhancement techniques have been applied to the same image, $I(x)$, where x is the pixel spatial location. Each enhancement technique produces an output $I_i(x)$ where $i \in [0, N)$ is a numerical reference to each of the enhancement techniques. Each enhancement technique has a subsequent quality metric matrix, $M_i(x)$, which contains the metric scores for each pixel. These matrices can be utilized to fuse all of the techniques together. This new enhanced image is given as

$$E(x) = \frac{1}{T_m(x)} \sum_{i=0}^{N-1} M_i(x)^p I_i(x), \quad (5.1)$$

where $T_m(x) = \sum_{i=0}^{N-1} M_i(x)^p$. The final enhanced image $E(x)$ is a local weighted sum of each of the enhancement techniques, based on the quality of each technique. p is typically a large value ($p \geq 35$) to allow for maximum separation between techniques of different resulting quality. Without this value, the individual weights do not mean as much and the output is more of a simple mean of the inputs instead of a smart weighted system. A large value of p will give a larger weight to a pixel value of the enhancement technique which performs best at spatial location x . Figure

5.2 gives a more in-depth look into the weighting system. The enhanced images first go through

Figure 5.2: Weighting Algorithm Flowchart

the no-reference metric described in Chapter III. However, instead of using the overall quality score, the score matrix which contains individual pixel scores is used. This matrix is put through a Gaussian low-pass filter to smooth the matrix before the fusion. Without the filter, the fused image has obvious transition points where the dominant enhancement technique changes. After filtering the score matrix, each element is raised to the p th power to get separation between the different techniques' scores. Finally, the single technique's score matrix is divided element-by-element with the total of all the techniques' score matrices. Once all of the techniques' matrices have gone through this process, they will each have their normalized weight on a per-pixel basis. The normalized score matrices are multiplied by the corresponding enhanced images and these are added up to create the final fused image.

CHAPTER VI

IMAGE FUSION TECHNIQUE RESULTS

To test the fusion technique, the original, GHE, CLAHE, and LTSN images are used to create an output fused image. The four input and one output image are all scored with the no-reference metric. Table 6.1 shows how the fused images compare to the originals and other enhanced images. This table shows that the fused image gets the highest score from the metric in every image. By

Table 6.1: No-reference Metric Scores for Original, Enhanced, and Fused Images

	Boat	Lakehouse	Door	Cave	Fireplace	Mountain House	Island
Original	0.6593	0.7896	0.5425	0.8140	0.6722	0.8553	0.8397
GHE	0.7666	0.7801	0.6803	0.7890	0.7609	0.8247	0.7636
CLAHE	0.7266	0.8276	0.6256	0.8718	0.7567	0.8684	0.8570
LTSN	0.7701	0.8135	0.5955	0.8375	0.7622	0.8266	0.8202
Fused	0.8115	0.8401	0.7049	0.8742	0.8031	0.8914	0.8698

weighting the enhancement techniques locally, a fused output is created that maximizes the best and minimized the worst of each technique. This fused output has statistics for lightness and mean closer to ideal than any of the images enhanced by a single technique.

Figures 6.1-6.14 show the images mentioned above in their original, enhanced, and fused forms.

Figure 6.1: Boat Images. Top left: Original, $M_s = 0.6593$, Top middle: GHE, $M_s = 0.7666$, Top right: CLAHE, $M_s = 0.7266$, Bottom left: LTSN, $M_s = 0.7701$, Bottom middle: Fused, $M_s = 0.8115$

Figures 6.1 and 6.2 show how the fused boat image compares to the original and enhanced versions. The fused image has the brightness and detail in the back of the boat that is present in GHE, but keeps the detail in the water present in LTSN and CLAHE.

Figures 6.3 and 6.4 are of the lakehouse image. The fine details in the trees that LTSN provides are preserved in the fused image. In Figure 6.4, a closer look at the actual house shows the GHE version of the lakehouse is darker than CLAHE and LTSN. The GHE image is also darker than the fused image, so LTSN and CLAHE must have had higher local scores in this area.

Figures 6.5 and 6.6 show the door image. Again, the fused image retains the brightness of GHE and the detail in the windows of the original, CLAHE, and LTSN. There are areas in the image where the transition points between dominating techniques are apparent. These can be mitigated by varying the low-pass filter the score matrices are put through, but the artifacts as shown do not hinder the information able to be seen in the image.

Figure 6.2: Boat Images. Top left: Original, Top middle: GHE, Top right: CLAHE, Bottom left: LTSN, Bottom middle: Fused

Figures 6.7 and 6.8 show the cave image. This image differs from the previous three in that the original is of already fairly high quality; that is, there are no extremely overexposed or underexposed areas. Still, the enhanced images show that the contrast can be improved. CLAHE does well on this image, so the fused output is very similar. However, there are some small areas where LTSN provides more detail, shown clearly in Figure 6.8. Although as a whole the LTSN image lacks in contrast, it illuminates the underside of the rocks well. This carries over to the fused image, which has the contrast present in CLAHE but the fine details shown in LTSN.

The fireplace image, shown in Figures 6.9 and 6.10, has both overexposed and underexposed areas. Details in the benches and fireplace itself are hard to see in the original. GHE and CLAHE brighten these areas, but overexposes the windows to an even greater degree. LTSN brightens the same areas but also does a good job in bringing out the patterns on the windows. These patterns show up in the fused image where the original and LTSN would have the highest weights in those areas. Also note the statue head, seen clearly in Figure 6.10. The fused image takes this portion from the GHE image. GHE and LTSN both illuminate this part of the image, but the noise in the LTSN version hurts its score and therefore its weight.

Figure 6.3: Lakehouse Images. Top left: Original, $M_s = 0.7896$, Top middle: GHE, $M_s = 0.7801$, Top right: CLAHE, $M_s = 0.8276$, Bottom left: LTSN, $M_s = 0.8135$, Bottom middle: Fused, $M_s = 0.8401$

Figures 6.11 and 6.12 show the mountain house image. Like the cave image, the original does not have any glaring problems. Still, the enhancement techniques can improve upon aspects of the image. The fused image has enhanced contrast that comes from GHE and CLAHE. LTSN shows up in the fused image in the form of detail in the fence and blinds.

Figures 6.13 and 6.14 show the island image, another that does not have many initial issues. The fused image takes the high contrast water from CLAHE and the detail in clouds from LTSN to create a better combined overall image.

Figure 6.4: Lakehouse Images. Top left: Original, Top middle: GHE, Top right: CLAHE, Bottom left: LTSN, Bottom middle: Fused

Figure 6.5: Door Images. Top left: Original, $M_s = 0.5425$, Top middle: GHE, $M_s = 0.6803$, Top right: CLAHE, $M_s = 0.6256$, Bottom left: LTSN, $M_s = 0.5955$, Bottom middle: Fused, $M_s = 0.7049$

Figure 6.6: Door Images. Top left: Original, Top middle: GHE, Top right: CLAHE, Bottom left: LTSN, Bottom middle: Fused

Figure 6.7: Cave Images. Top left: Original, $M_s = 0.8140$, Top middle: GHE, $M_s = 0.7890$, Top right: CLAHE, $M_s = 0.8718$, Bottom left: LTSN, $M_s = 0.8375$, Bottom middle: Fused, $M_s = 0.8742$

Figure 6.8: Cave Images. Top left: Original, Top middle: GHE, Top right: CLAHE, Bottom left: LTSN, Bottom middle: Fused

Figure 6.9: Fireplace Images. Top left: Original, $M_s = 0.6722$, Top middle: GHE, $M_s = 0.7609$, Top right: CLAHE, $M_s = 0.7567$, Bottom left: LTSN, $M_s = 0.7622$, Bottom middle: Fused, $M_s = 0.8031$

Figure 6.10: Fireplace Images. Top left: Original, Top middle: GHE, Top right: CLAHE, Bottom left: LTSN, Bottom middle: Fused

Figure 6.11: Mountain House Images. Top left: Original, $M_s = 0.8553$, Top middle: GHE, $M_s = 0.8247$, Top right: CLAHE, $M_s = 0.8684$, Bottom left: LTSN, $M_s = 0.8266$, Bottom middle: Fused, $M_s = 0.8914$

Figure 6.12: Mountain House Images. Top left: Original, Top Middle: GHE, Top right: CLAHE, Bottom left: LTSN, Bottom middle: Fused

Figure 6.13: Island Images. Top left: Original, $M_s = 0.8397$, Top middle: GHE, $M_s = 0.7636$, Top right: CLAHE, $M_s = 0.8570$, Bottom left: LTSN, $M_s = 0.8202$, Bottom middle: Fused, $M_s = 0.8698$

Figure 6.14: Island Images. Top left: Original, Top Middle: GHE, Top right: CLAHE, Bottom left: LTSN, Bottom middle: Fused

CHAPTER VII

CONCLUSIONS AND FUTURE WORK

7.1 Conclusion

Image quality is important in many fields within the image processing community. Images are used to store and convey information, and a high quality image ensures the most information is passed on. A no-reference image enhancement quality metric has been proposed to evaluate image enhancement methods. Historically, these methods have been difficult to evaluate objectively. Many of the existing image quality metrics require a reference image and/or make assumptions about distortions present in the imagery. The proposed metric uses statistics to locally score images based on lightness, contrast, and noise content. The idea of locally scoring the images is continued in the proposed image fusion technique, where the original and enhanced images are fused together in a way that maximizes the good and minimizes the bad of each individual method. It is shown that this fused output image scores higher than any of the others using the proposed no-reference metric.

7.2 Future Work

Determining image quality without a reference is an important and current problem in the image processing community. While researching other quality methods, a “blind” no-reference quality

metric was discovered [11]. More in-depth research will be done into this method and how it compares to the method proposed in this paper.

The robustness of the proposed system can be tested by adding and/or removing enhancement techniques. Currently tests have been done with GHE, CLAHE, and LTSN, but there are multitudes of enhancement techniques that could be tested as well. Also, work can be done on accelerating the metric scoring system and the fusion technique both in hardware and software. Furthermore, research can be done on how to expand this system to color imagery. Some enhancement methods (such as LTSN) are designed to work with color images. Statistics could be extracted from the intensity (luminance) of a color image, but fusing multiple color images together may be a less trivial task.

BIBLIOGRAPHY

- [1] R. C. Gonzalez and R. E. Woods, *Digital Image Processing (3rd Edition)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006.
- [2] B. Girod, “What’s wrong with mean-squared error?” in *Digital Images and Human Vision*, A. B. Watson, Ed. Cambridge, MA: MIT Press, 1993.
- [3] H. R. Sheikh and A. C. Bovik, “Image information and visual quality,” *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, Feb. 2006.
- [4] Z. Wang, H. R. Sheikh, and A. Bovik, “No-reference perceptual quality assessment of JPEG compressed images,” in *Image Processing. 2002. Proceedings. 2002 International Conference on*, vol. 1, 2002, pp. I–477–I–480.
- [5] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld, “Adaptive histogram equalization and its variations,” *Computer Vision, Graphics, and Image Processing*, vol. 39, no. 3, pp. 355 – 368, 1987.
- [6] S. Arigela and V. Asari, “Visibility improvement of aerial imagery by a locally tuned nonlinear enhancement technique,” in *Image Analysis and Interpretation (SSIAI), 2012 IEEE Southwest Symposium on*, April 2012, pp. 217–220.
- [7] D. J. Jobson, Z. Rahman, and G. A. Woodell, “Statistics of visual representation,” in *Proc. SPIE*, vol. 4736, 2002, pp. 25–35.
- [8] Z. Rahman, D. J. Jobson, and G. A. Woodell, “Retinex processing for automatic image enhancement,” pp. 390–401, 2002.
- [9] D. L. Donoho and I. M. Johnstone, “Adapting to unknown smoothness via wavelet shrinkage,” *Journal of the American Statistical Association*, pp. 1200–1224, 1995.
- [10] E. Balster, Y. Zheng, and R. Ewing, “Feature-based wavelet shrinkage algorithm for image denoising,” *Image Processing, IEEE Transactions on*, vol. 14, no. 12, pp. 2024–2039, Dec 2005.
- [11] A. Mittal, R. Soundararajan, and A. Bovik, “Making a completely blind image quality analyzer,” *Signal Processing Letters, IEEE*, vol. 22, no. 3, pp. 209–212, March 2013.