Robust Defocus Blur Identification in the Context of Blind Image Quality Assessment

Izak van Zyl Marais, Willem Herman Steyn
University of Stellenbosch, Stellenbosch 7600, South Africa
20th May 2009

Abstract

A defocus blur metric for use in blind image quality assessment is proposed. Blind image deconvolution methods are used to determine the metric. Existing direct deconvolution methods based on the cepstrum, bicepstrum and on a spectral subtraction technique are compared across 210 images. A variation of the spectral subtraction method, based on a power spectrum surface of revolution, is proposed and is found to compare favourably with existing direct deconvolution methods for defocus blur identification. The method is found to be especially useful when distinguishing between in-focus and out-of-focus images.

1 Introduction

1.1 Context of the Problem

The work presented here is part of the development of an onboard image processing system for a low earth orbit, remote sensing micro satellite. Since downlink bandwidth and time are limiting factors in the satellite image acquisition chain, onboard image processing is often used to manage down-link data requirements [1, 2]. In our application this processing takes the form of prioritising images for download according to a quality metric. Given that the amount of images that can be downloaded in a day is less than the amount that can be acquired, it is desirable to download only the the acquired images with the best quality.

Remote sensing images can be degraded in various ways: cloud cover can obscure targets, sensor noise can corrupt images, an unstable satellite can cause geometric distortions and smearing and a defocused lens can cause blur. We attempt to measure some of these degradations and combine their effect into a single quality score for each image. Therefore, we are working on an image quality assessment problem. Although

*Corresponding Author. E-mail addresses: imarais@sun.ac.za, whsteyn@sun.ac.za
the focus of this paper is on the estimation one feature of the quality assessment algorithm (namely defocus extent), it is presented in the context of autonomous blind image quality assessment.

Sections 1.2 and 1.3 give short overviews of the fields of image quality assessment and blur estimation. In section 2 the theoretical base of the class of blur estimation algorithms central to this paper is discussed. Section 3 justifies a comparative test of the techniques presented in section 2. Angular spectral smoothing, a new technique based on the method discussed in 2.4, is presented section 4. Finally, the methods from sections 2 and 4 are compared in section 5 and conclusions are drawn in section 6.

1.2 Image Quality Assessment

The goal of image quality assessment is to enable a machine to make an objective judgement on the quality of an image that corresponds to a subjective evaluation of the same image by humans. The majority of existing literature focuses on full-reference quality assessment algorithms, where the original image is assumed to be available for comparison with the degraded image [3, 4, 5, 6]. The need for a reference image limits the application of these algorithms and differentiates them from the humans, who can easily determine the quality of an image without a reference. This has led to the formulation of the blind image quality assessment (also called no-reference) problem, in which an attempt is made to appraise the quality of an image without reference.

Applications discussed for blind image quality assessment include, among others, intelligent memory management in digital cameras [7] and measuring the performance of both contrast enhancing algorithms [8] and compression algorithms [9]. Since we also do not have access to reference images, we must use blind image quality assessment.

Li [7] proposes three quantities as objective measures to aid in blind image quality assessment: edge sharpness level, random noise level and structural noise level. Although Li makes no such distinction, it is important to separate image degradation measures and image content measures. Random and structural noise are degradation measures, whilst edge sharpness level is a content measure. Degradation measures are generally more objective than content measures: an image with low SNR is almost always worse than an image with high SNR. The same cannot be said for content measures: in remote sensing there is little reason to believe images containing more sharp edges in the global structure are more useful than images containing little or none. However, temperature variations can cause the telescope to become defocused, potentially leading to a useless set of images that contain no sharp edges. The problem is therefore to distinguish between images where the subject has little or no sharp edges and images blurred by the imaging system. Such images might achieve the same score using an edge sharpness level measure. Indeed, this problem resulted in outliers in the image quality assessment experiments in [9], where blind use of an edge sharpness metric caused some images to receive disproportionately bad scores.

Hence, for our image quality assessment we measure only degradation measures: the defocus blur extent of the point spread function (PSF) is chosen instead of edge sharpness level. As part of a subjective study we conducted, 20 viewers evaluated 140
remote sensing images degraded with varying levels of defocus blur extent. The results in figure 1 are comparable to the Gaussian blur results from [3] and confirm a monotonic relationship between perceived image quality and defocus extent.

Figure 1: Difference mean opinion scores for 140 degraded images. Larger scores correspond to worse perceived quality.

1.3 Blur Estimation

Blind estimation of the PSF is a subset of the blind image deconvolution problem, which, given the linear degradation model

\[ g(x, y) = f(x, y) * h(x, y) + n(x, y), \]  

attempts to recover the original image, \( f(x, y) \), and PSF, \( h(x, y) \), using only the degraded image, \( g(x, y) \), and partial information about the imaging system. \( n(x, y) \) models additive noise (white, signal independent Gaussian is commonly assumed) and \(*\) is the two dimensional convolution operator. Since the problem is ill-conditioned and may lead to non-unique solutions, many approaches have been suggested. These cover a broad range in terms computational complexity and applicability, with some tailored to specific scene types and others to specific PSF’s. An instructive overview can be found in [10, 11]. Although many techniques are discussed, they follow one of two approaches:

1. Identify the PSF first and then use a classical technique such as Wiener filtering to restore the image. This approach is called \textit{a priori} or \textit{direct} blur identification. Algorithms in this category are computationally simple.

2. Identify the PSF and true image simultaneously. An iterative process is used that estimates the PSF, restores the image, evaluates the result and then repeats. Many algorithms fall into this \textit{indirect} category. They are generally computationally complex and often have ill-convergence.
Recent advances postdating [10] include an autocorrelation based direct method for motion blur identification capable of identifying both linear and accelerated motion [12, 13, 14]. If access to multiple instances of the same image blurred by substantially different PSF’s is available, [15] can be used in either direct or indirect configuration. An alternative approach to direct identification uses vector quantisation to train a classification system to recognise various types of PSF, but requires the system be trained for specific images [16]. A direct method that allows identification of complicated, non-linear motion PSF paths is [17], which uses special hardware to achieve high spatial and temporal resolution.

For our purposes, the class of direct methods based on spectral techniques was chosen [18, 19, 20]. Despite recent advances discussed above and in [10], these methods remain popular because they make no assumptions about the true image structure and are computationally simple. These attributes make them suitable for an onboard implementation in a remote sensing system, since earth images do not conform to a template and processing of very high resolution images using limited memory and processing power is required.

2 Blur Identification Based on Spectral Techniques

2.1 Degraded Image Model

Since no constraints are placed upon the true image, assumptions about the blur PSF shape are required to make the blind deconvolution problem solvable. Defocus blur is modelled by a uniform PSF with 2-D circular support with radius $R$:

$$
 h(x, y) = \begin{cases} 
 0; & \sqrt{x^2 + y^2} > R \\
 \frac{1}{\pi R^2}; & \sqrt{x^2 + y^2} \leq R. 
\end{cases} 
$$

(2)

2.2 Power Spectrum and Power Cepstrum

Cannon proposed the use of the power cepstrum for blur identification [18]. If we first consider power spectra, (1) becomes:

$$
 P_g(u, v) = P_f(u, v) |H(u, v)|^2 + P_n(u, v). 
$$

(3)

$H(u, v)$, the frequency response of the PSF, is of the form $J_1(Rr)/(Rr)$ where $R$ is the PSF radius, $r = \sqrt{u^2 + v^2}$ and $J_1(\cdot)$ is the first-order Bessel function of the first kind (which has nearly periodic, radial zero-crossings). Welch’s method [21] is used to reduce the variance of the $P_g$ estimation:

$$
 P_g(u, v) = \frac{1}{N} \sum_{i=1}^{N} |G_i(u, v)|^2
$$

(4)

The image is subdivided into square sections. Each section, $g_i(x, y)$, is windowed and the power spectra of all sections are averaged to estimate $P_g(u, v)$. The radial zero-crossings
of \( H(u, v) \) are zeros in \(|H(u, v)|^2 \) and local minima in \( P_g(u, v) \). Blur identification in the spectral domain proceeds by identifying the first local minimum.

If we consider instead the power cepstrum\(^1\),

\[
C_g(p, q) = \mathcal{F}^{-1}\{\log P_g(u, v)\},
\]

where \( \mathcal{F}^{-1} \) is the inverse Fourier transform, the defocus blur radius is characterised by a ring of large negative spikes at \( 2R \) from the origin in \( C_g(p, q) \). These are assumed to be the result of periodic zeros in \( P_g(u, v) \). During blur identification only the negative part \( C_g(p, q) \) is considered. Using the cepstrum instead of the spectrum has some advantages: it is algorithmically easier to identify a global negative maximum rather than a first significant local minimum. Furthermore, since the ring of spikes is the result of periodic minima in the spectrum, it is more robust to noise than identification based on only one minimum.

### 2.3 Bispectrum and Bicepstrum

In an attempt to increase the reliability of blur identification in the presence of Gaussian noise Chang et. al. [19] turned to the bispectrum [23]. Since the bispectrum is the Fourier transform of the third-order moment sequence (which is zero for a stationary, zero-mean Gaussian process), the bispectrum is invariant to Gaussian noise. Chang et. al. suggests that for blur identification it is sufficient to consider only the 2-D “central slice” or 1-D “central-line”, which can be computed more efficiently than the full 4-D bispectrum. The recommended direct 2-D estimator for the \( i \)th sub-segment is

\[
B_g^{(i)}(u, v; 0, 0) = G_i(u, v)G_i(0, 0)G_i^*(u, v).
\]

The mean of the entire observed image has to be removed before it is segmented. In a similar fashion to (4), averaging is used to reduce the variance:

\[
B_g(u, v; 0, 0) = \frac{1}{N} \sum_{i=1}^{N} B_g^{(i)}(u, v; 0, 0).
\]

Since defocus blur has circular symmetry, the local minima in the 1-D \( B_g(u, 0; 0, 0) \) give the same information as in the 2-D \( B_g(u, v; 0, 0) \) and are used for identification.

Savakis and Easton [20] rely for identification on negative peaks in the bicepstrum:

\[
D_g(p, q) = \mathcal{F}^{-1}\{\log B_g(u, v; 0, 0)\}.
\]

They argue that the use of bicepstrum over bispectrum holds the same advantages as cepstrum over spectrum, but inherits robustness against noise from the bispectrum.

---

\(^1\)Some definitions use the forward transform instead [22]. They and are functionally equivalent.
2.4 Spectral Subtraction and Comb Filtering

As an alternative to higher order spectra, Fabian and Malah [24] proposed adding pre- and post-processing to the cepstral method to increase robustness in the presence of noise. The pre-processing is based on a spectral subtraction technique [25], which attempts to estimate the spectrum of the blurred, noiseless image, \( A(u, v) = F(u, v)H(u, v) \), by subtracting a weighed estimate of the noise power spectrum from the degraded image power spectrum:

\[
|\hat{A}(u, v)| = \begin{cases} \sqrt{P_g(u, v) - \alpha \hat{P}_n(u, v)} & \text{if } P_g(u, v) > \alpha \hat{P}_n(u, v) \\ \epsilon & \text{otherwise} \end{cases}
\]

where \( \epsilon \) is a small constant used to avoid numerical issues when taking the logarithm. This approach rests upon the idea that, for white Gaussian noise, \( P_n \) is a constant offset proportional to the noise variance. This offset “obscures the zeros” [24]: when taking the cepstrum, it hinders the logarithm’s ability to accentuate local minima in the power spectrum. Since the logarithm is nonlinear, the degree to which minima are accentuated is highly dependant on their proximity to zero. A median-complement filtered image is used as an approximation of the noise-image from which \( \hat{P}_n(u, v) \) is computed. Fabian and Malah claim, and this was confirmed by our tests, that the method gives better results if the image is not subdivided.

\( \hat{A}(u, v) \) is used to compute the cepstrum \( C_a(p, q) \), which has circular symmetry. Although this implies using the most negative peak in \( C_a(p, 0) \) is sufficient for blur identification, the variance caused by noise makes such an approach unreliable. Instead, an angular average \( C_a(r) \) is computed by converting \( C_a(p, q) \) to polar coordinates \( C_a(r, \theta) \) and averaging over \( \theta \).

Aside from the main pulse at \( 2R \), \( C_a(r) \) exhibits harmonics at values of \( r \) approximately multiples of \( 2R \). In the presence of noise there are also spurious peaks at other values which may dominate the true peak at \( 2R \). The post-processing step employs an adaptive comb-like filter that amplifies peaks which have harmonics (like the true peak) and suppresses peaks which do not have harmonics (like the spurious peaks). The filter is:

\[
C_l(r) = \frac{|C_a(r)|}{\sqrt{M \sum_{i \in A_r} (C_a(i))^2}}
\]

for quefrency \( r \), \( A_r = \{i | i > r_0 \text{ and } i \notin (kr - 1, kr, kr + 1), k = 0, 1, 2 \ldots \} \).

\( A_r \) is the “disturbance set”: the set of quefrencies where harmonics of \( r \) are not expected. This set resembles a comb-filter with 3-point stop bands. \( M \) is the total number of points in \( A_r \). \( r_0 = 3 \) to avoid an \( A_r \) consisting only of stop bands, which would be an empty set. The output of the filter is therefore limited to values of \( r > 3 \).
3 Problems with Methods in Existing Literature

3.1 Lack of Comparative Defocus Tests

In spite of popularity of direct methods there exists little or no comparative literature on the subject. Although [14] compares the cepstral method with their method for images with a variety of PSF blur extents and SNRs, their method is only applicable to motion blur. Reference [20] compares the methods of [19] and [18] with their own and apply only the post-processing from [24] (in our experience it is the pre-processing that is responsible for most of the method’s performance). Their comparison is based on degraded images generated form a single test image and only motion blur identification is tested, in spite of the fact that defocus blur is more difficult to identify [24].

Ability to operate at low SNR is commonly used as a measure of algorithm performance, with comparison to other methods often based on best achieved SNR:

\[
SNR = 10 \log \left( \frac{\sigma_s^2}{\sigma_n^2} \right) \tag{12}
\]

We have found that, for this class of methods, SNR is a poor indicator of performance. Although, for a given image, the SNR is highly correlated with the identification capability (increasing noise variance \(\sigma_n^2\) or decreasing signal variance \(\sigma_s^2\) have a detrimental effect), when comparing different images \(\sigma_s^2\) does not play as important a role as signal frequency content. This is because, the more the high frequency content in \(P_f(u, v)\), the better the periodic zeros at higher frequencies in \(|H(u, v)|^2\) are visible in \(P_g(u, v)\). For example, we could correctly identify blur in a city scene, Figure 2(d), which has dense spatial activity and good high frequency content, up to SNR as low as 2.8 dB. Using a desert scene, 2(a), with sparse spatial structure, 17 dB was the best that could be achieved. This confirms the need for a comparative test across a variety of image types, since comparing methods based on best reported SNR is of little use. The results of such a test are presented in section 5.

3.2 Inappropriate Generalisation from 1-D to 2-D

Both [19] and [20] use 1-D, 256 pixel strip image sections for the averaging in (7). They only give results for motion blur, which, if sections are taken along the blur direction, is a 1-D function. However, they suggest that 1-D sections are equivalent to 2-D square image sections and it is implied that their results are valid for defocus blur as well. Our experiments show that this is not the case. We define classification error distance:

\[
e_d = \frac{|R - \hat{R}|}{R}, \tag{13}
\]

where \(\hat{R}\) is the estimated defocus blur extent. Using the five test images in Figure 2, defocus blur was added using blur extents \(R = \{2, 3, 4, 9, 15\}\), resulting in 25 blurred images. Since no noise was added, the cepstral method could be used. 1-D sections of 256 pixels and 2-D sections of 128×128 pixels were both used and \(e_d\) was compared,
Figure 2: Base images used during experiment, with their resolutions.
with results in Table 1. It is clear that when identifying a 2-D PSF, using 2-D sections have a significant advantage.

<table>
<thead>
<tr>
<th>Sections</th>
<th>Images with $e_d &lt; 10%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-D</td>
<td>13 (52%)</td>
</tr>
<tr>
<td>2-D</td>
<td>25 (100%)</td>
</tr>
</tbody>
</table>

Table 1: Comparison of defocus blur classification accuracy using 1D and 2D image sections.

As mentioned in section 2.4, [24] uses a $3 \times 1$ median-complement filter to estimate $\hat{P}_n(u, v)$ for spectral subtraction. However, we found that the filter imposes an unwanted structure on the power spectrum visible in Figure 3(b); it does not estimate the Gaussian noise as white, but instead concentrates noise energy at high frequencies. Therefore, it cannot “uncover the zeros” present in noiseless, blurred power spectrum, 3(a), at low frequencies. If the $\hat{P}_n(u, v)$ estimate is averaged over all frequencies prior to subtraction, the location of the first low-frequency zero can be uncovered 3(c), greatly improving classification accuracy.

That power spectra similar to 3(c) can be used in cepstral blur identification is surprising, since the cepstral peak has been previously assumed to result from radially periodic zeros of the $J_1(Rr)/(Rr)$ function [18, 24, 20]. Although periodic zeros are a requirement for a cepstral peak when using 1-D image strips, it was found that, when using the 2-D cepstrum, periodicity is not a requirement. It does however increase the relative height and accuracy of the peak. This was confirmed in an experiment on a blurred image where the periodicity of the power spectrum was removed. First the angular average of the first significant local radial minimum was computed. Then all power spectrum content at radial frequencies greater than the first local minimum was set equal to this average. Figure 4 shows the effect of this non-linear low-pass-filter on the power spectrum. Taking the cepstrum of this filtered power spectrum still allowed
identification of the defocus blur extent in spite of no periodicity in the spectral domain (Figure 4). The resulting peak was however slightly shifted (10-15%).

Figure 4: Effect of removing radial periodicity in $P_g(u,v)$ on $C_g(p,q)$. Note that only the zero-clipped negative part of $C_g(p,q)$ is shown: white indicates large negative values and black indicates zero.

This observation strengthens the case for using 2-D sections when identifying defocus blur. Consequently, 2-D sections are used in the angular smoothing method in section 4 as well as the comparison test in section 5.

4 Angular Spectral Smoothing

4.1 Avoiding Power Spectrum Distortion

The clipping element of the spectral subtraction technique is inherent in equation (9). However this clipping of negative values to $\epsilon$ effectively distorts the shape of the power spectrum: the average of radial minima is increased relative to the rest of the signal. This effect can be seen in Figure 5(a). This figure was made by converting $P_g(u,v)$

Figure 5: Power spectra. (a) shows the distortion resulting from clipping during spectral subtraction. (b) shows the effect of angular smoothing prior to subtraction.
into polar coordinates $P_g(r, \theta)$ and averaging over $\theta$ to give $P_g(r)$. The original blurred, noiseless power spectrum $|A(r)|^2$, the noisy power spectrum $P_g(r)$, and the estimate by spectral subtraction $|\hat{A}(r)|^2$, are shown. The distortion at the first local minimum is clearly visible.

To avoid this effect, one could decrease the subtraction extent to $(\min(P_g(u, v)) - \epsilon)$. However, even with the use of Welch’s method, the variance in $P_g(u, v)$ means that such a restriction severely limits the amount of subtraction, and therefore noise mitigation, possible. Instead we attempt further to reduce the variance of $P_g(u, v)$ prior to subtraction.

4.2 Smoothing Procedure

To achieve this the same cartesian to polar conversion used to generate the plots in Figure 5, is used as a starting point. First $P_g(u, v)$ is estimated according to (4). It is converted to $P_g(r, \theta)$ using bilinear interpolation. $P_g(r, \theta)$ is averaged to $P_g(r)$ to reduce variance. $P_g(r)$ values for $r > N/2$, where $N \times N$ is the size in pixels of the sections used, are set equal to the average of $P_g(N/2)$, $P_g(N/2 - 1)$ and $P_g(N/2 - 2)$. This is done because values beyond $r = N/2$ map to the corners of $P_g(u, v)$. These values are therefore estimated over fewer angles, making the averaging less reliable. The higher frequencies are also dominated by noise power which further increases their variance.

Next, the 1-D $P_g(r)$ is used as a profile to create a surface of revolution: the 1-D sequence is swept around the origin, $r = 0$, of the 2-D space in the $\theta$ direction. This process creates a 2-D power spectrum $P_g(r, \theta)$, which is angular smoothed: all pixels at the same radius, say $r = k$, have the same value, namely the value of the angular average at radius $r = k$ in $P_g(r, \theta)$. $P_g(r, \theta)$ is converted back to cartesian $P_g(u, v)$, again using bilinear interpolation. The process is illustrated in Figure 6. While the averaging to 1-D reduces variance, use of a surface of revolution enforces circular symmetry on the power spectrum. The inherent circular symmetry of the defocus blur power spectrum is therefore strengthened while features of the image power spectrum $P_f(u, v)$ are further suppressed. This suppression can be seen in Figure 6: the diagonal stripes visible in 6(a) are image features and are not present in 6(b). Following angular smoothing, the spectral offset created by $P_n(u, v)$ can be maximally removed in a manner similar to (9), but without the clipping:

$$\hat{P}_a(u, v) = |\hat{A}(u, v)|^2 = P_g(u, v) - \min\{P_g(u, v)\} + \epsilon.$$ (14)

Figure 5(b) shows the amount of distortionless spectral subtraction made possible by this technique. Note also that no noise estimate $\hat{P}_a(u, v)$ is required. Although replacing the median filter, used to estimate $\hat{P}_a(u, v)$ in the spectral subtraction technique, with a more sophisticated estimator might also improve results, a comparative test has shown that no noise estimator is capable of consistently accurate estimates across a wide range of image types and noise levels [26]. Thus, removing the need for noise estimation increases the robustness of our method.

The value of $\epsilon$ should ideally be as small as possible to maximise noise mitigation. However, we found that, especially for noiseless images, choosing $\epsilon$ too small dispropor-
Figure 6: Use of angular smoothing to reduce variance and enforce circular symmetry. Note that, for illustration purposes, $P_g(u, v)$ was estimated from the whole image and not using Welch’s method.

Tionately accentuates the regions of $\overline{P}_g(u, v)$ closest to zero when taking the logarithm. These regions are typically at high frequencies and are not the local minima that we wish to accentuate. Whether a certain numerical value of $\epsilon$ should be considered low or high is a function of the block size used during Welch’s method. We found, that for $128 \times 128$ pixel blocks, $10 < \epsilon < 100$ gives good results. At these levels the spectral offset is still lower than would result from gaussian noise with $\sigma_n = 0.01$.

The cepstrum $C_a(p, q)$ is taken from $\hat{P}_a(u, v)$ and clipped so that only negative values are considered. Except for quantisation effects this exhibits perfect circular symmetry, so to get $C_a(r)$ a central slice can be used. The post-processing filter from (11) can be used on this 1-D sequence to enhance the height of the desired peak. A simple peak picking algorithm identifies the blur radius.

An unexpected advantage of this angular smoothing approach is that in-focus images have a peak at $r = 2$ that can be detected prior to comb-filtering $C_a(r)$ (as previously explained, the output of the filter is restricted to $r > 3$). Since image power spectra are generally exponential in shape, the surface of revolution created from a profile in which the corner frequencies are set equal to a constant, is a circular shape with the first radial local minimum at the edge of $\overline{P}_g(u, v)$. This reliably maps to a cepstral sequence similar to the one in Figure 9 and is a boon in the context of blind image quality assessment, since in-focus images can be easily identified.

4.3 Estimate Confidence

Any deconvolution technique will fail given a low enough SNR and concerns have been expressed about the sensibility of using blind (and hence unreliable) image quality as-
essment in a scientific environment [6]. Therefore, during blur identification, it is useful to have an indication of the confidence in the blur estimate; in a fully automated, scientific environment images should only receive a bad score if it can be stated with certainty that they are out of focus. $C_a(r)$ will always have some maximum. At low noise this maximum is at a prominent peak $2R$ from the origin. However, as SNR decreases, spurious peaks appear and eventually dominate $C_a(r)$. The concept of relative energy, $E_r$, is introduced to enable differentiation between true and spurious peaks:

$$E_r = \frac{E_{\text{peak}}}{E_{\text{rest}}}$$

$$E_{\text{peak}} = \sum_{i \in P} (C_a(i))^2,$$

$$P = \{r | r_p - 1 \leq r \leq r_p + 1\},$$

$$r_p = \arg\max [C_a(r)],$$

$$E_{\text{rest}} = \sum_{i \in Q} (C_a(i))^2,$$

$$Q = \{r | r \notin P, 0 \leq r \leq r_{\text{max}}\}.$$

The relative energy in the peak of the cepstral sequence $C_a(r)$ is computed. It can be tested against a threshold relative energy level and, if it is too low, the identification is probably wrong and the results should not be trusted, otherwise the defocus blur estimate is $\hat{R} = r_p/2$. For a image with $R = 4$ ($r_p = 8$), Figure 7 shows the effect of adding noise with an increasing standard deviation, $\sigma_n$, on $E_r$. 7(b) and 7(c) show $C_a(r)$: the prominent, correct peak in (b) results in high $E_r$, while spurious peaks in (c) result in low $E_r$.

<table>
<thead>
<tr>
<th>$\sigma_n$</th>
<th>$E_r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>206</td>
</tr>
<tr>
<td>1</td>
<td>125</td>
</tr>
<tr>
<td>3</td>
<td>6.56</td>
</tr>
<tr>
<td>5</td>
<td>6.48</td>
</tr>
<tr>
<td>7</td>
<td>0.662</td>
</tr>
<tr>
<td>9</td>
<td>0.349</td>
</tr>
<tr>
<td>20</td>
<td>0.334</td>
</tr>
</tbody>
</table>

Figure 7: Spurious peaks dominate at higher $\sigma_n$ and result in lower $E_r$.

When using the original cepstral method or the biceptral method, low $E_r$ is also typical of images that have no blur. Distinguishing in-focus and out-of-focus images are important for blind image quality assessment. This is discussed further in section 5.
5 Comparative Experiment

5.1 Experimental Set-up

To evaluate the performance of the angular smoothing method and compare the existing spectral based, direct methods, an experiment was done on a range of images. The 5 base images from Figure 2 were chosen, since they represent a wide variety of remote sensing image types. (a), (c) and (e) have low spatial detail and varying degrees of edge gradient levels, (b) has sharply defined edges typical of coastal regions and (d) has high spatial detail typical of city scenes. All images are 8-bit greyscale with resolutions shown in the figure. These images where blurred with defocus radii $R = \{0, 2, 3, 4, 9, 15\}$ and white, zero mean, Gaussian noise was added with standard deviations $\sigma_n = \{0, 1, 3, 5, 7, 9, 20\}$, resulting in a total of 210 images.

The cepstral, bicepstral, spectral subtraction and angular smoothing methods where all used to identify the defocus blur radius. For the spectral subtraction method, $P_g(u, v)$ was estimated from the whole image and $\alpha = 1$ was used, as recommended in [24]. For the other methods, $128 \times 128$ pixel sections were averaged according to (4) and (7) to estimate $P_g(u, v)$ and $B_g(u, v; 0, 0)$ respectively. In all cases the resulting 2-D (bi)cepstral sequence was averaged to 1-D, to facilitate peak picking. The comb-filter post-processing was applied in all cases except for the basic cepstral technique. After peak picking, $E_r$ thresholds for each method were varied from minimum to maximum using deciles (10 equally spaced $E_r$ indices were chosen from a sorted list of output $E_r$ values). Based on these $E_r$ thresholds images were separated into “classified” and “unclassified” groupings, illustrated in Figure 8 for the angular smoothing method. The number of classified images is not zero for maximum $E_r$, because in-focus images were classified differently.

![Figure 8: Effect of varying $E_r$ on number of classification.](image)

Images that were possibly in focus were classified using special rules, since none of
the methods (barring spectral subtraction, discussed below) gives \( r_p = 0 \) for in-focus images. Since the comb-filtering technique zeros all values for \( r < 4 \), identification of in-focus images was done prior to comb-filtering. Given an in-focus image, cepstral sequences for the different methods are shown in Figure 9. The bicepstrum typically gives results similar to the cepstrum for in-focus images: neither method gives a peak at a characteristic location, but the \( E_r \) level is normally low. As already discussed, the angular smoothing method results in a peak at \( r = 2 \). Images processed using the cepstral, bicestral and angular smoothing methods with \( r_p \leq 2 \) were assumed to be in focus irrespective of \( E_r \) level. In our experience, the spectral subtraction method has a large peak at \( r = 0 \) for all image types (both in-focus and blurred). This can be explained by looking at Figure 3(c): clipping a large part of the power spectrum to the same \( \epsilon < 1 \) value results in a large low frequency (quefrency) component in \( C_a(p, q) \) that is negative, since \( \log(\epsilon) < 0 \). Therefore, identification by spectral subtraction could only be done after comb-filtering, which restricts the range of the output cepstral sequence to \( 4 \leq r \leq r_{\max} \). Consequently, images processed with the spectral subtraction method were assumed in-focus if \( r_p = 4 \).

For each classification, the error distance was calculated according to (13) and averaged across all classifications for a given method. If \( R = 0 \) and \( \hat{R} \neq 0 \), \( e_d = 100\% \), which corresponds to an in-focus image, incorrectly classified as out-of-focus. Since number of classifications vary with \( E_r \) threshold, methods are compared using average \( e_d \) against number of classified images.

5.2 Results

Figure 10 shows the result of the comparative test. The usefulness of the \( E_r \) measure is obvious from the fact that the classification accuracy generally increases with decreasing number of classifications. First consider the right side of the graph where all 210 images are classified. The results confirm the high noise sensitivity of the cepstral method to
additive noise; the many noisy images present result in a large average $e_d$ of 232%. The bicepstral method shows a big improvement with average $e_d$ of 71%, while the spectral subtraction and angular smoothing give $e_d = 36\%$ and $e_d = 29\%$ respectively. As the number of classifications decrease, $e_d$ generally improves, except in the case of spectral subtraction. The increase in this case is caused by the inability of the method to distinguish between in-focus images and images blurred with $R = 2$ ($r_p = 4$), which is not affected by increasing the $E_r$ threshold. Instead these incorrectly classified images with $e_d = 100\%$ just start to make up a bigger portion of the total classifications, resulting in an increase in average $e_d$. Many in-focus images were also classified as having $r_p = 5$, 6 or 7 and generally the method has lower accuracy when used with small blur radii.

Depending on the application, the rules for in-focus classification and the use of $R = 2$ as one of the defocus blur radii might seem to tip the advantage unfairly in the direction of angular averaging. As an additional test, all-in focus images and images classified as in-focus were discarded and the same plots generated. Figure 11 shows the results. The maximum number of classifications achieved with the spectral subtraction technique is in this case fewer than the others, since more images had to be discarded. The angular spectral smoothing still compares favourably.

These results confirm the usefulness of the angular smoothing technique, especially in the context of blind image quality assessment.
6 Conclusions

Across 210 test images, the angular smoothing method gave the best average classification error of all the direct, spectral based, defocus blur identification methods evaluated. The increased robustness of the method can be ascribed to the fact that no noise estimate is needed, as well as the reinforcement of the circular structure typical of defocus blur. The method’s characteristic response to in-focus images represents a further improvement. This makes it especially suitable for use in situations where examined images may be either in-focus or out-of-focus, such as during blind image quality assessment. The usefulness of the relative energy threshold was proven for situations where certainty about the estimate accuracy is required.

Acknowledgments

I gratefully acknowledge the funds received from the Wilhelm Frank bursary fund.

References


