A NO-REFERENCE SPATIAL ALIASING MEASURE FOR DIGITAL IMAGE RESIZING

Amy R. Reibman
AT&T Labs – Research
Florham Park, NJ 07932

Shan Suthaharan*
Department of Computer Science
The University of North Carolina at Greensboro
Greensboro, NC 27402

ABSTRACT

We present a no-reference Spatial Aliasing Measure (SAM) for images, that assesses the impact of visual jaggy edges in resized digital images. The measure makes use of the fact that visible spatial aliasing is likely to occur in areas with strong directional energy. In these regions, aliasing energy often appears in frequency regions that have little image content. We show the effectiveness of SAM using a subjective test.

Index Terms— Image quality, aliasing, jagginess, image resizing, image scaling

1. INTRODUCTION

The aliasing introduced during image resizing or wavelet coding degrades image quality by causing pixellation (or jagginess), geometric deformation, and inhomogeneity in contrast [1]. Digital image content is widespread today, so that even the naive viewer is experienced in seeing and identifying these jaggy artifacts. Recently, spatial aliasing caused by image resizing has been explored, both in terms of the design of polyphase filters to reduce aliasing [2] and with the introduction of a family of filters that provide implementers a simple way to choose filters that optimize the trade-off between blurring and aliasing during image resizing [3]. However, these do not consider the visual impact of aliasing.

Methods to measure image quality are improving rapidly. Perceptual models consider low- and mid-level aspects of the human visual system to quantify the impact of image degradation (see for example [4], which considers the low-level perceived contrast and mid-level global precedence properties). These are usually full-reference (FR) metrics which rely on the availability of an original image for comparison. Performance among a number of FR metrics are compared in [5].

However, in many applications, an original image is not available. No-reference (NR) image quality metrics are typically feature-based models, which focus on measuring only the impact of one or a few specific artifact types. (See for example the comparison among NR blocking and blurring metrics in [6].) These individual impairments metrics can also be combined to characterize the combined visual impact of multiple specific artifacts. For example, Farias and Mitra [7] present a NR metric that combines blockiness, blurriness, and noisiness for video compression. Image resizing has an inherent trade-off between blurring, ringing, and aliasing. NR metrics exist for the first two artifacts [8]; however, none exist for aliasing artifacts.

Vicario et al. [1] present a methodology to assess image quality after resizing based on 5 types of visual artifacts: blurring, ringing, and the three aliasing artifacts mentioned above. They use expert viewers to judge the severity of each of these impairment types to obtain an overall quality value for a resized image. However, the use of expert viewers severely limits the applicability of their method.

While there is no feature-based metric that considers aliasing, Daly presents an insightful study of FR quality assessment for aliasing in downsampled images [9]. In particular, he shows that horizontal and vertical edges will mask any aliasing; however, edges not at exactly 0, 45, or 90 degrees will produce visible artifacts because the frequency support of the aliasing does not coincide with that of the signal. He concludes that a FR metric that measures the strength of aliasing independent of its orientation will not be accurate.

In this paper, we present a no-reference measure which measures the impact of aliasing for image resizing, with an emphasis on image down-sampling. Our measure focuses on the aliasing energy associated with visible jagginess in areas with strong directional energy. We specifically consider only those aliasing contributions which are not masked by the corresponding signal energy. Section 2 describes the impact of aliasing. Section 3 describes our NR measure, and performance results are shown in Section 4.

2. IMPACT OF SPATIAL ALIASING

In this section, we describe the impact of aliasing in image resizing. We begin by writing an expression for the aliasing signal for image resizing, and then show how the aliasing affects both the picture quality and the image spectrum. We present the equations using one dimension for notational simplicity; the extension to two dimensions is straightforward.

Figure 1 shows an architecture for image resizing. The input image \( x[n] \) is upsampled by a factor \( L \), filtered by an
aliasing-reduction filter $h[n]$, and then downsampled by a factor $M$ to create the output image $y[n]$. The Discrete-Time Fourier Transform (DTFT) of the output image is

$$Y(\omega) = \frac{1}{M} \sum_{k=0}^{M} Y_I\left(\frac{\omega}{M} + \frac{2\pi k}{M}\right)$$

$$= \frac{1}{M} Y_I\left(\frac{\omega}{M}\right) + \frac{1}{M} \sum_{k=1}^{M} Y_I\left(\frac{\omega}{M} + \frac{2\pi k}{M}\right)$$

where $Y_I(\omega) = H(\omega)X_I(\omega)$, $H(\omega)$ is the DTFT of $h[n]$, and $X_I(\omega) = X(L\omega)$ is the DTFT of the intermediate signal prior to filtering. The sum in (1) is the aliasing due to down-sampling. The folding process described by (1) adds high-frequency content from $x[n]$ into the low frequencies of $y[n]$, if these frequencies are not completely eliminated by $H(\omega)$.

However, the first term may also contain aliasing if up-sampling occurs, i.e. if $L > M$. The content of $X(\omega)$ from $\pi/L \leq |\omega| < \pi$ will produce aliasing unless attenuated by the filter $H(\omega)$. Therefore, we define $Y_0(\omega)$ to be the DTFT of the resized signal that would be obtained if the filter perfectly removes all content with $|\omega| \geq \pi/\max(M,L)$. Then the actual aliasing in $Y(\omega)$, using non-ideal $H(\omega)$, is

$$A(\omega) = Y(\omega) - Y_0(\omega)$$

$$= H\left(\frac{\omega}{M}\right) X_0\left(\frac{L\omega}{M}\right) + \frac{1}{M} \sum_{k=1}^{M} H\left(\frac{\omega}{M} + \frac{2\pi k}{M}\right) X\left(\frac{L\omega}{M} + \frac{2\pi kL}{M}\right)$$

where $X_0(L\omega/M) = 0$ for $|\omega| < \pi/\max(M,L)$ and $X(L\omega/M)$ otherwise.

Figure 2 illustrates the impact of aliasing on an image. Figure 2(a) shows the full Harbor image that has been downsampled by a factor of $M = 4$ after filtering with a non-ideal filter. To analyze the aliasing, we decompose the image into overlapping patches, each windowed by a Blackman-Harris window [10] to eliminate image-boundary effects. Figure 2(b) shows one such (unwindowed) patch from the lower-left corner of this image, at actual size. Aliasing is clearly visible in the near-horizontal line in the middle of the patch. For comparison, Figure 2(c) shows the identical patch without aliasing. Figure 2(d) shows the spectrum of Figure 2(b) (log of the squared-magnitude of the DFT with the DC value removed) with the spectrum from Figure 2(c) in Figure 2(e).

---

**Fig. 1.** Resizing by a rational factor of $L/M$.  

**Fig. 2.** Harbor image. (a) Heavily-aliased image; (b) patch from (a); (c) same patch without aliasing; (d) spectrum of (b); (e) spectrum of (c); (f) directional and aliasing mask for nearly-vertical energy; (g) 89 selected patches.

In both Figures 2(d) and (e), we see two areas of strong directional energy, emanating from the DC location in the center. These energies are perpendicular to the strong edges in the center of the image patch. However, in Figure 2(d) we also see one strong aliasing component which was introduced during downsampling by the spectral-folding process described in equation (1). This aliasing component does not go through the origin, but instead is a folded-over continuation of the directional energy. Figure 2(f) presents a binary mask that illustrates the location of directional energy and its folded aliasing for the first term in the summation of (1), for the nearly-horizontal directional energy in Figure 2(d).
For this image patch, this folded energy is easily visible in the downsampled spectrum because it is added into a frequency region that has little energy in the original image patch. It is this observation that forms the basis for our no-reference aliasing measure described next.

3. NO-REFERENCE SPATIAL ALIASING MEASURE

It is inherently difficult to assess image degradation in a no-reference environment, because it is difficult to distinguish between artifacts and the actual desired image content. However, as noted above, in image patches in which most of the energy is primarily directional (but not at 0, 45, or 90 degrees), it is likely that the folded aliasing energy will have little overlap with the signal content. Therefore, in these image patches, it should be possible to partition the overall energy into aliasing and signal components.

Our no-reference aliasing measure has four main stages. First, we identify image patches likely to contain strong directional energy with few distractions. Second, for each patch, we estimate a finely-quantized direction for this energy. Third, for each patch, we partition the total energy into signal energy that will mask any aliasing energy. Thus, aliasing should be measured only for those frequencies in M, which is the extension of the directional mask Ms into the downsampled spectrum because it is added into a reference aliasing measure described next.

Stage 1: Identify patches with strong directional energy.
We would like to quickly identify areas with strong directional energy, areas that are not too busy but have edges. Define the local contrast \( C_y(m, n) = \frac{\sigma_y(m, n)}{\mu_y(m, n)} \) where \( \mu_y(m, n) \) is the local mean and \( \sigma_y(m, n) \) is the local standard deviation of the image \( y \) over a 7 × 7 pixel region centered on pixel \((m, n)\). Regions with strong local contrast may contain strong edges, but they may also be very busy. We compute \( \sigma^2_y(m, n) \), the local variance of this contrast image at pixel \((m, n)\). Regions with small \( \sigma^2_y(m, n) \) but large \( C_y(m, n) \) are likely to have many edges. Regions with a large contrast variance, however, are more likely to be strong isolated edges.

Thus, we search for the maximum contrast variance, and center our first patch. We then restrict any subsequent patches from being within \( N = 16 \) pixels from the selected patch, and search for the next maximum. We continue this process until we have selected 20% of the available patches. Figure 2(g) illustrates the set of patches, \( P \), chosen using this method on the Harbor image; patches are shown multiplied by a Blackman-Harris window centered on the patch center.

Stage 2: For each patch, find a precise estimate of the direction.
The goal of our detection algorithm is to find the main directional energy that might cause aliasing, without becoming distracted by low-frequency energy or by any strong aliasing components in the high and middle frequencies. For each patch identified in the previous part, we subtract its mean and apply a Blackman-Harris window to eliminate image-boundary effects. We desire a precise estimate of the direction of the energy, so that we can pinpoint the aliasing component in the frequency domain. We search for energy in all possible directions using masks created using Bresenham lines [11].

Let \( Z^p(i, j) \) be the squared magnitude of the 2-D DFT of the \( p \times p \) windowed patch, for \(-p/2 \leq i, j < p/2\). Let \( M_s(i, j) \) define a directional mask in the frequency domain from \((-s, -p/2 + 1)\) to \((s, p/2 - 1)\), for \(-p/2 < s < p/2\). (Horizontal masks are similarly defined, and denoted by index \( s + p \)). \( M_s(i, j) \) is a binary line of width \( w = 5 \) created using Bresenham’s algorithm [11]. We similarly create an aliasing mask \( A_s(i, j) \) which is the extension of the directional mask \( M_s(i, j) \) folded (once) as if during downsampling. The sum of \( A_s(i, j) \) and \( M_s(i, j) \) is shown in Figure 2(f) for one value of \( s \). Our direction-finding algorithm follows.

Step 1: Initialize \( k = 1 \) and initialize \( Q \), the set of frequencies to consider, to be \( Q = \{(i, j) | 3 \leq |i|, |j| \leq \alpha(p/2)\} \). Initialize \( \beta = \sum_{(i, j) \in Q} Z^p(i, j) \).

Step 2: For each \( s \), compute the energy in that direction:
\[
d_s = \sum_{(i, j) \in Q} M_s(i, j) Z^p(i, j) \quad (3)
\]
If \( \max_s d_s < \beta \epsilon \), set \( k_{\text{max}} = k - 1 \) and stop. Otherwise, save \( s(k) \) to be the \( s \) which maximizes \( d_s \).

Step 3: Remove frequencies in \( M_s(i, j) \) and \( A_s(i, j) \) from \( Q \).

Step 4: Set \( k = k + 1 \) and go to Step 2 if \( k < 4 \).
We set \( \alpha = 0.4 \) and \( \beta = 0.05 \).

Stage 3: For each patch, estimate signal energy without aliasing components.
In this stage, we estimate the signal energy in frequencies that are identified as containing aliasing components, based on the directions \( s(k) \) above. First, we form two sets, \( M^p \) and \( A^p \), the unions over all \( k \leq k_{\text{max}} \) of the masks \( M_s(k) \) and the aliasing masks \( A_s(k) \), respectively. Frequencies in \( M^p \) are known to have strong signal energy that will mask any aliasing energy. Thus, aliasing energy should be measured only for those frequencies in \( A^p \) that are not in \( M^p \).

We begin by estimating the isotropic energy, \( Z^p_{\text{iso}}(i, j) \), that remains after removing the strong directional components in \( M^p \). This forms a lower bound on the estimated signal energy for the frequencies in \( A^p \) but not in \( M^p \). An upper bound on the estimated signal energy is the observed \( Z^p(i, j) \). We estimate the signal energy \( Z^p_{\text{sig}}(i, j) = \min(Z(i, j), \max(Z^p(i, j), Z^p_{\text{iso}}(i, j))) \), where \( Z^p(i, j) \) is the average energy in \( Z^p(i, j) \) on either side of the mask \( A_s(k)(i, j) \), for those frequencies in \( A^p \) but not in \( M^p \).

Stage 4: Estimate the severity of the aliasing error.
We compute the Signal-to-Aliasing Ratio (SAR) across all measured patches as
\[
\text{SAR} = 10 \log_{10} \left( \frac{\sum_{(i, j) \in Q} Z^p(i, j) - \sum_{(i, j) \in P} Z^p(i, j)}{\sum_{(i, j) \in Q} Z^p(i, j)} \right) \quad (4)
\]
Scoreboard
aliased
middle left image. Viewers strongly disliked the two most
they preferred the image from the middle of Table 1(b), and
the image with the most blur (top right of each subtable in
aliasing to those with more. For Expt. 2, viewers always rated
but otherwise preferred the downsampled images with less
which has some limited aliasing for
per original image; viewers in Expt. 2 made 28.
Tive subjective quality within the set of degraded images, for
experiment using paired comparison, and their responses are
first rows of Table 1(b) produce no aliasing.
ω
larger
Kω
transition widths
and ringing artifacts. For example, for the same
Kω
transition width
has flat passband and stopband with cut-off frequency
degraded images are created using a raised cosine filter which
ω
five degraded images per original are produced using an ideal
Images:
Twenty-two viewers rated the degraded images in each
For each degraded image, we compute our SAM and
Marziliano’s blur and ringing metric [8], relative to the
ideally-downsampled image. A_i, B_i, and R_i are the relative
aliasing, blur, and ringing, respectively. We apply
stepwise regression on these parameters to predict relative
quality q_i, and obtain the combined models
q_i = -7.562 0.006321 R_i − 1.070 A_i and
q_i = -0.2777 0.002311 R_i for Expts. 1 and 2, respectively. Pearson’s, Spearman’s and
Kendall’s correlation are 0.85, 0.65, 0.47 for Expt. 1 and 0.64,
0.47, and 0.36 for Expt. 2, respectively.
Conclusions: We present a novel technique to measure
the spatial aliasing in a resized digital image without access to
the original image. We combine our aliasing metric with
existing ringing and blurring metrics using linear regression
to characterize subjective image quality.

5. REFERENCES

’05, March 2005.
1965.

<table>
<thead>
<tr>
<th>M = 2</th>
<th>0.50π</th>
<th>0.55π</th>
<th>0.60π</th>
<th>0.65π</th>
<th>0.70π</th>
</tr>
</thead>
<tbody>
<tr>
<td>M = 4</td>
<td>0.25π</td>
<td>0.30π</td>
<td>0.35π</td>
<td>0.40π</td>
<td>0.45π</td>
</tr>
</tbody>
</table>

\( Kω \)

<table>
<thead>
<tr>
<th>M = 2</th>
<th>0.50π</th>
<th>0.41π</th>
<th>0.33π</th>
</tr>
</thead>
<tbody>
<tr>
<td>M = 4</td>
<td>0.25π</td>
<td>0.208π</td>
<td>0.166π</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Kω</th>
<th>0</th>
<th>0.2π</th>
<th>0.5π</th>
</tr>
</thead>
<tbody>
<tr>
<td>M = 2</td>
<td>0.50π</td>
<td>0.41π</td>
<td>0.33π</td>
</tr>
<tr>
<td>M = 4</td>
<td>0.25π</td>
<td>0.208π</td>
<td>0.166π</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>0.528π</td>
<td>—</td>
</tr>
</tbody>
</table>

Table 1. Parameters for creating degraded test images. Cut-off frequency \( ω_0 \) for (a) Expt. 1, ideal filter. (b) Expt. 2 raised cosine filter for \( Kω \). \( M = 2 \) Images: Flowchart and City. \( M = 4 \) Images: Harbor, Bridge, Scoreboard, Restaurant.

where the numerator is the estimated energy in the signal
without aliasing, and the denominator is the estimated energy in
the aliasing components.

4. PERFORMANCE RESULTS

We conducted two subjective tests using six original images
downsampled with \( L = 1 \) and \( M = 2 \) or \( M = 4 \). In Expt. 1,
five degraded images per original are produced using an ideal
filter with different cut-off frequencies \( ω_0 \). In Expt. 2, eight
degraded images are created using a raised cosine filter which
has flat passband and stopband with cut-off frequency \( ω_0 \) and
transition width \( Kω_0 \) [12]. The filter parameters, detailed in
Table 1, are chosen to produce a range of aliasing, blurring,
and ringing artifacts. For example, for the same \( ω_0 \), shorter
transition widths \( Kω \) produce more ringing and less aliasing;
larger \( ω_0 \) produces blurring but reduces aliasing. Filters in the
first rows of Table 1(b) produce no aliasing.

Twenty-two viewers rated the degraded images in each
experiment using paired comparison, and their responses are
analyzed using the Bradley-Terry model [13] to predict relative
subjective quality within the set of degraded images, for
each original image. Viewers in Expt. 1 made 10 comparisons
per original image; viewers in Expt. 2 made 28.

For Expt. 1, viewers preferred the second image in the set
which has some limited aliasing for Bridge and Flowchart,
but otherwise preferred the downsampled images with less
aliasing to those with more. For Expt. 2, viewers always rated
the image with the most blur (top right of each subtable in
Table 1(b)) as having the lowest quality, although the image
they preferred among the set was not consistent. Three times
they preferred the image from the middle of Table 1(b), and
once each the ideally filtered image, the middle top, and the
middle left image. Viewers strongly disliked the two most
aliased Scoreboard images in Expt. 1.

For each degraded image, we compute our SAM and
Marziliano’s blur and ringing metric [8], relative to the