L_0-Regularized Image Downscaling

Junjie Liu, Shengfeng He, *Member, IEEE,* and Rynson W.H. Lau, *Senior Member, IEEE*

***Abstract*—In this paper, we propose a novel L_0-regularized optimization framework for image downscaling. The optimization is driven by two L_0-regularized priors. The first prior, gradient-ratio prior, is based on the observation that the number of edges in the downscaled image is approximately inverse square proportional to the downscaling factor. By introducing L_0 norm sparsity to the gradient ratio, the downscaled image is able to preserve the most salient edges as well as the visual perception of the original image. The second prior, downsampling prior, is to constrain the downsampling matrix so that pixels of the downscaled image are estimated according to those optimal neighboring pixels. Extensive experiments on the Urban100 and BSDS500 datasets show that the proposed algorithm achieves superior performance over the state-of-the-arts, in terms of both quality and robustness.**

***Index Terms*—Image downscaling, L_0 norm sparsity, salient edges preserving**

I. Introduction

Image downscaling is arguably the most fundamental operation used on mobile phones and the Web for fast browsing purposes. In contrast to image super-resolution, which aims to produce a higher resolution image by *"creating"* image content, the goal of image downscaling is to produce a lower resolution image by *"preserving"* the original image content as much as possible.

Traditional downscaling algorithms mainly address the aliasing problem. They first filter the image to blur the edges and then subsample it to produce the downscaled image [14]. While this helps reduce the aliasing problem, the image is usually over-smoothed and the fine details cannot be retained, as the filter kernel is independent of the image content. Kopf *et al.* [10] address this problem by estimating the kernel shapes to align adaptively with the local image content in a bilateral manner. However, these filter-based methods generally require a certain minimum kernel size to be effective, and fail to preserve small scale details or repeated patterns (see the first example in Fig. 1c). An alternative approach is to downscale the image without filtering [16], which constructs the downscaled image by directly optimizing the SSIM index [23] between the original and downscaled images. Although this approach can capture most of the perceptually important details, some of the irrelevant details, e.g., noise, are exaggerated (see the second example in Fig. 1d). In addition, due to the spatial-independent nature of SSIM, the optimization process may lead to jagged edge artifacts in structured regions, and color splitting artifacts may also occur as SSIM can only be optimized on individual channels (see the car grilles in the first example of Fig. 1d).

In this paper, we propose a L_0 optimization framework for image downscaling. Our optimization focuses on two critical issues of image downscaling, *salient features preservation* and *downscaled image construction*. Accordingly, two L_0-regularized priors are proposed. The first prior, gradient-ratio prior, is proposed based on the observation that the number of edge pixels of the downscaled image is inverse square proportional to the downscaling factor. This prior is able to constrain the downscaled image to preserve the most salient edges and visual perception of the original image. The second prior, downsampling prior, aims to control the sparsity of the downsampling matrix, i.e., the number of neighboring pixels that are integrated to form each downscaled pixel value, to select the best pixel combination. More importantly, the downsampling prior allows us to formulate the downscaling problem as a sparse coding problem, so that the downscaled image can be directly optimized with the guidance of the original image, avoiding the undesirable artifacts. The proposed L_0-regularized optimization can be solved iteratively by splitting into two sub-problems, while each sub-problem converges very fast as they have close-form solutions.

Extensive qualitative and quantitative evaluations illustrate that the proposed method outperforms the state-of-the-arts in terms of perceptual quality and robustness. Specifically, the proposed method preserves salient features well in structured regions and regions with repeated patterns, without ringing nor exaggeration artifacts.

II. Related Work

Image downscaling algorithms can be categorized into two main groups as follows.

Content preservation methods: aim to construct the original image in a lower resolution without losing image details. Traditional algorithms involve three steps: filtering, subsampling, and construction. Different filter kernels, e.g., Lanczos and bicubic filters [14], are proposed to suppress the high frequency components in order to avoid aliasing artifacts. However, these filter-based algorithms introduce undesirable ringing or over-smoothing artifacts. Recently, some methods [15], [21] propose to apply interpolation on image downscaling by adding an extra construction step to reduce these artifacts, but they still suffer from blurred fine details.

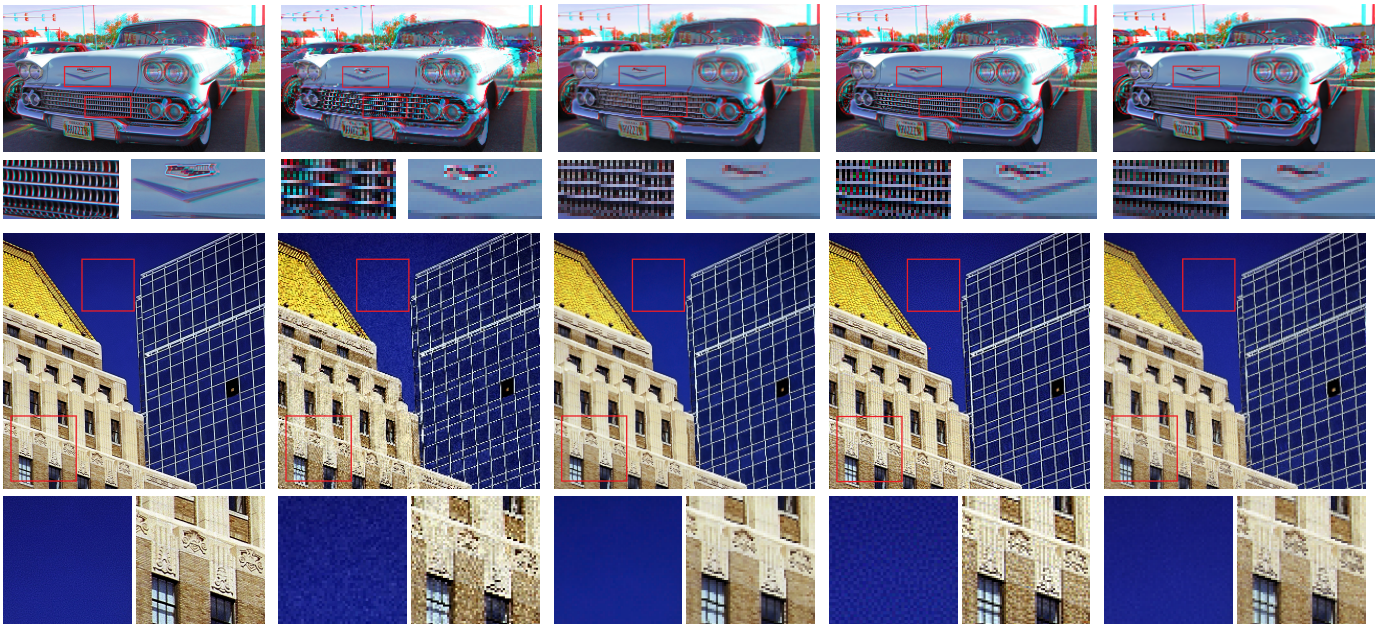
An alternative filter-based method [10] adaptively estimates the filter kernel by considering both spatial and color variances

This work was supported in part by the Natural Science Foundation of China under Grant 61702194, and two SRG grants from City University of Hong Kong (No. 7004676 and 7004889). (Corresponding author: Shengfeng He.)

Junjie Liu is with the School of Computer Science and Engineering, South China University of Technology, China, and the Department of Computer Science, City University of Hong Kong, Hong Kong. E-mail: junjie.liu.cs@gmail.com.

Shengfeng He is with the School of Computer Science and Engineering, South China University of Technology, Guangzhou, China. E-mail: hesfe@scut.edu.cn.

Rynson W.H. Lau is with the Department of Computer Science, City University of Hong Kong, Hong Kong. E-mail: Rynson.Lau@cityu.edu.hk.



(a) Original (b) Subsampling (c) Content-adaptive [10] (d) Perceptual based [16] (e) Ours

Fig. 1: Limitation of existing downscaling methods. While the state-of-the-art filter-based method [10] (c) may not be able to capture small details or cause jagged edge artifacts, the perceptual based method [16] (d) may exaggerate noise and split the colors. In contrast, the proposed L_0-regularized algorithm is able to preserve most of the salient edges with little artifacts.

of the local region. This helps improve the crispness of the downscaled images, especially for pixel-art images. However, filter-based methods are restricted by the kernel size and do not take visual perception into account. As a result, important details are not well preserved. Instead of finding a better filter kernel, a recent method tries to optimize the downscaled images [16]. It proposes a SSIM-based [11] optimization framework to preserve the visual perception of the original image. Taking advantages of the SSIM metrics in a global optimization, [16] is able to render most of fine details. However, as SSIM is spatial-independent and the optimization ignores the spatial distribution of pixels within each patch, noises and jagged edges are amplified. In contrast, we optimize the downscaled images according to salient edge variation, which helps preserve edges and prevent amplifying artifacts.

Salient object guided methods: (also called image retargeting/resizing methods) aim to preserve the salient objects in the downscaled images [5]. They are typically used in situations where the downscaled image has a different aspect ratio than the original, and their main objective is to maintain the aspect ratio of the salient objects while reducing the size of the image through wrapping or seam carving [2], [17]. Saliency guided cropping [20], [18] is also used to produce thumbnails [12], so as to highlight the representative objects. The proposed method belongs to the content preservation category, which has a very different objective to salient object guided methods.

III. L_0-Regularized Priors for Image Downscaling

To compute a downscaled image that faithfully retains the original perception, we aim to find out "*what*" the most effective features are to preserve and "*how*" to construct the low resolution image. In this section, we propose two L_0-regularized priors, gradient-ratio and downsampling priors, to address these two problems. We also propose a new optimization framework to iteratively regularize two downscaled images, the ratio-preserved image x and the downsampling image y_l, to obtain the final result in x.

A. The Objective Function

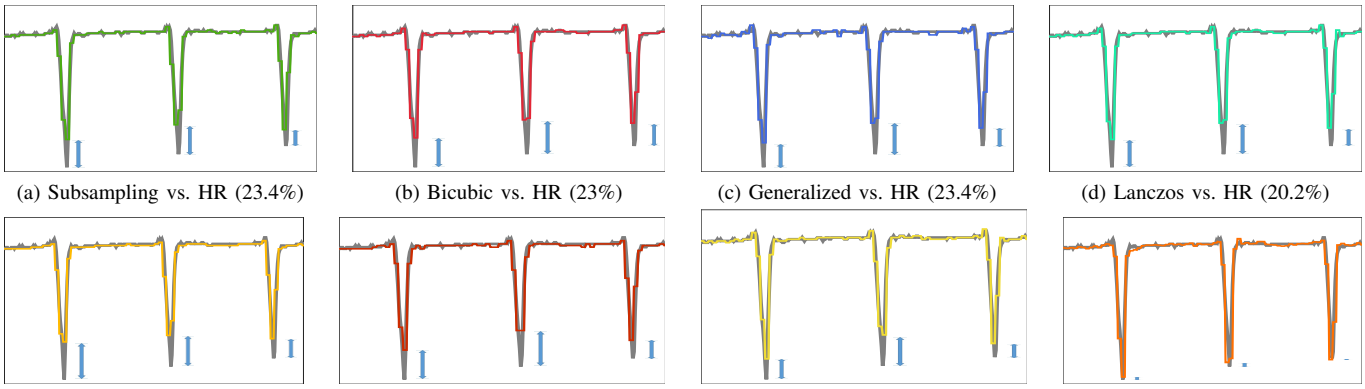
Before presenting the proposed priors, we first introduce the objective function of the image downscaling problem. Given an input image y, our objective function is defined as:

$$\min_{x,m} ||x - y_l(m)||_2^2 + \sigma|\mu P(\nabla x) - P(\nabla y)| + \lambda P(m), \quad (1)$$

where x and $y_l(m)$ are the two downscaled images needed to be optimized. These two images are iteratively optimized by fixing one image to solve another. x is the downscaled result under the constraint of the gradient-ratio prior, while $y_l(m)$ is an downscaled image generated by downsampling matrix m under the downsampling prior. $P(\cdot)$ is the L_0-regularized operator. σ, λ and μ are the parameters to constrain the optimization function, and will be elaborated later when we discuss the corresponding sub-objective functions. The proposed optimization framework allows the input and output images the have different sizes by introducing the downsampling prior for cross-scale optimization. On the other hand, the proposed gradient-ratio prior constrains the number of edge pixels in the downscaled image in order to maintain a similar visual perception as the original one. In the following subsections, we divide this objective function to two subproblems with the proposed priors, which are iteratively optimized.

B. L_0-Regularized Gradient-ratio Prior

As demonstrated by experiments and evidences in cognitive psychology [22], [19], edges play a very important role in



(a) Subsampling vs. HR (23.4%) (b) Bicubic vs. HR (23%) (c) Generalized vs. HR (23.4%) (d) Lanczos vs. HR (20.2%)

(e) Bilateral vs. HR (29.2%) (f) Content-adap. [10] vs. HR (23%) (g) Percept. based [16] vs. HR (15%) (h) Ours vs. HR (6%)

Fig. 2: Comparison of different downscaling methods on a 1D scanline of a HR image. State-of-the-art methods cannot capture salient edges, while the proposed method reproduces the original signal well, especially on fine details. We show the relative average differences of the three "v" spikes (blue arrows) in brackets. The downscaled scanlines are upscaled to the same size as the original ones by pixel replication for illustration. (Best viewed at full size on a high-resolution display).

human visual perception. Early human visual processing operates by perceptually grouping edges into object-like structures for preliminary recognition [3]. As one of the main visual perception indicators, edges have been used to maintain or manipulate image visual perception in different applications, such as image filtering [7], [8] and detail enhancement [6]. They (in particular those salient edges) are one of the most important properties to be preserved in image downscaling. However, all existing image downscaling methods do not consider this factor. Fig. 2 shows a 1D scanline of an example high resolution (HR) image, and the downscaled scanlines (followed by upscaling with pixel replication) from different downscaling methods. All existing methods fail to preserve the original intensity of salient edges (Fig. 2b-g). (See Section V-D for a thorough evaluation on edge preservation.) Our goal here is to preserve the intensity of salient edges and control the number of edge pixels, which are critical in preserving the original visual perception.

To constrain the number of edge pixels, we introduce L_0 norm on image gradient. The number of non-zero gradient edge pixels of an image a is defined as:

$$P(\nabla a) = ||\nabla a||_0, \tag{2}$$

where $||\nabla a||_0$ indicates the number of non-zero gradient edge pixels of a. This discrete counting function is then used as a constraint to confine the downscaled image. Given an original image y and its ideal downsampling version y_l, our L_0-regularized gradient-ratio prior can be used to optimize the downscaled image x to preserve the salient edges. The objective function of the gradient-ratio prior is defined as:

$$\min_x ||x - y_l(m)||_2^2 + \sigma|\mu P(\nabla x) - P(\nabla y)|, \tag{3}$$

where μ is the square of the downscaling factor, and σ is a weight to balance the two terms. This objective function aims to preserve the intensity (first term) and the number of edge pixels (second term) in the downscaled image. However, this function requires the guidance of an downscaled image y_l, as x cannot be optimized with the guidance image of different sizes. Obviously, an ideal downscaled input image is not available. Hence, we propose a downsampling prior to compute y_l in the optimization, so that it can preserve the content of y as faithfully as possible.

C. L_0-Regularized Downsampling Prior

We aim to compute a downsampling matrix m to indicate the weight of each pixel in the original image for constructing the downscaled version. These weights are linearly combined with the original pixel values within each local patch to obtain a final downscaled value. However, as the product of y and m does not give a patch-based correlation, we present a new way of correlating y and m by rearranging the pixel/weight positions in a patch-based manner. The main idea of this rearrangement is illustrated in an example in Fig. 3, and the downsampling image y_l can be determined by:

$$y_l(m) = diag(y_p \times m_p), \tag{4}$$

where $diag(s)$ extracts the diagonal elements from matrix s. m_p and y_p are rearranged from m and y. While each patch of y is rearranged as a row vector to form y_p, each patch of m is rearranged as a column vector to form m_p. As shown in Fig. 3, the size of y is 6×6 and the downscaling factor is 2. We first generate a downsampling matrix m with the same size as y. The patch size and patch step depend on the downscaling factor. In this example, the patch size is set to 2 and there is no overlapping between patches. Finally, there are 9 patches in the image. By rearranging each patch to a row vector in y_p, the height of y_p is equal to the number of patches in y and the width is equal to the number of pixels in one patch. Similarly, the width of m_p is equal to the number of patches in m and the height is equal to the number of pixels in one patch. According to Fig. 3, the dimension of y_p is 9×4 and of m_p is 4×9. With this rearrangement, each patch is now correlated to the diagonal elements of the resulting matrix of $y_p \times m_r$. y_l is obtained by selecting the diagonal values.

Since each pixel in y_l can be determined by only a few neighboring pixels, most of the pixels in y are redundant.

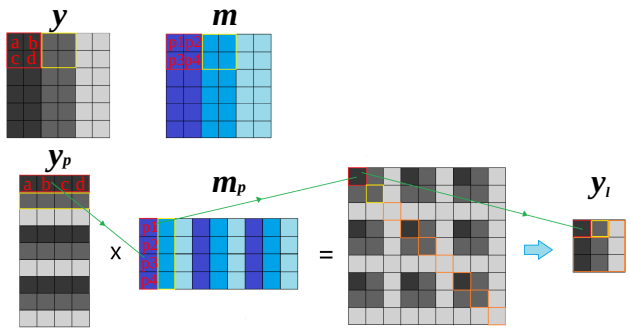


Fig. 3: (Top) The input HR image y and the downsampling matrix m. (Bottom) y and m are rearranged to matrices y_p and m_p, such that each local patch (2×2 pixels in the red box) of y is transformed to a row vector in y_p, and the patch in m is rearranged to a column vector in m_p. All patches are non-overlapping with each other. The downsampling image y_l is obtained by extracting the diagonal values from $y_p \times m_p$.

Hence, we introduce another L_0-regularized term to constrain the number of non-zero elements in m as:

$$P(m) = ||m||_0. \tag{5}$$

The objective function of the downsampling prior is then defined as:

$$\min_m ||x - y_l(m)||_2^2 + \lambda P(m), \tag{6}$$

where λ is the weight to balance the two terms. Thus, this downsampling prior ensures that m is sparse. Like filter-based methods, each pixel in the downscaled image is determined by the weighted sum of the neighboring pixels. However, unlike these methods, it is not restricted by the kernel size and shape (and the number of pixels), and m is optimized to the local image content. More importantly, our approach integrates filtering and subsampling into one step, which is guided by the original image instead of an image quality assessment metric (e.g., SSIM). As a result, the proposed construction method does not lead to over-smoothing nor exaggeration.

IV. Iterative Optimization

Our method uses two different optimization algorithms to solve the objective function (Eq. 1) iteratively. In general, it takes about 5 iterations to converged, and the ratio-preserved image x is selected as the final result.

A. Estimating x with m

Given an initial m (initialized as a bicubic filter), we first estimate x in here. As Eq. 3 involves a L_0 regularization term w.r.t. to the pixel-wise difference, it is difficult to find a close-form solution. Similar to [25], we use the half-quadratic splitting scheme as an efficient approximation. We rewrite the objective function of Eq. 3 by introducing an auxiliary variable $g = (g_h, g_v)^\top$, corresponding to the horizontal/vertical differences of x, as:

$$\min_{x,g} ||x - y_l(m)||_2^2 + \beta||\nabla x - g||_2^2 + \sigma|\mu P(g) - P(\nabla y)|, \tag{7}$$

where β is an adapting parameter to control the similarity between g and its gradient. When β is close to ∞, Eq. 7 approaches to Eq. 3. Variables x and g in Eq. 7 can be solved alternatively by fixing the other one. We first set g to zero, and the solution of x can be obtained by minimizing:

$$\min_x ||x - y_l(m)||_2^2 + \beta||\nabla x - g||_2^2. \tag{8}$$

This function is quadratic, and its closed-form solution can be efficiently computed as:

$$x = \mathscr{F}^{-1}\left(\frac{\mathscr{F}(y_l) + \beta(\overline{\mathscr{F}(\partial_h)}\mathscr{F}(g_h) + \overline{\mathscr{F}(\partial_v)}\mathscr{F}(g_v))}{\mathscr{F}(1) + \beta(\overline{\mathscr{F}(\partial_h)}\mathscr{F}(g_h) + \overline{\mathscr{F}(\partial_v)}\mathscr{F}(g_v))}\right), \tag{9}$$

where $\mathscr{F}$ and $\mathscr{F}^{-1}$ are the Fast Fourier Transform (FFT) and inverse FFT operators. $\overline{\mathscr{F}()}$ denotes the complex conjugate operator. ∂_h and ∂_v correspond to the derivative operators in the horizontal and vertical directions.

Once x is computed, g can be computed by minimizing:

$$\min_g \beta||\nabla x - g||_2^2 + \sigma|\mu P(g) - P(y)|. \tag{10}$$

The solution of g can then be obtained as:

$$g^x = \begin{cases} \nabla x, & \nabla x \geqslant \frac{\sigma\mu}{\beta} \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

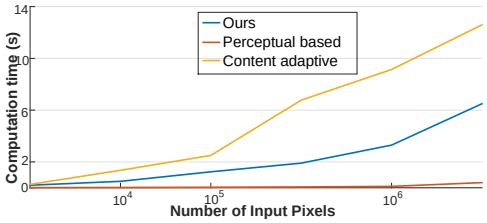
Algorithm 1 Solving Eq. 12.

Input: Original image y_p, the downscaled image x, predefined integer ξ to constrain the sparse elements, and the allowed value of maximal residual to converge τ;
Output: Downsampling image y_l and updated m_p;
Initialization: Initialize m_p;
for $i = 0; i \leq n$ **do**
  $f_0 \leftarrow x_i, k \leftarrow 0$;
  while $k \leqslant \xi$ **and** $f_{k+1} \leqslant \tau$ **do**
    $f_{k+1} = \arg\min_{m_p^i} ||y_p^i \times m_p^i - f_k||$;
    $k++$;
  end
end
Normalize m_p;

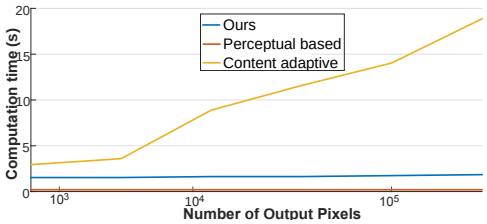
B. Estimating m with x

Given a downscaled image x, our goal here is to find a downsampling matrix m so that x can be represented by the diagonal values of the linear combination of original image y and m. This means that we aims to find the minimum number of high resolution elements to represent the downscaled image. This process can be viewed as a sparse coding problem, where y_p can be considered as a ground truth overcomplete dictionary and m_p the corresponding sparse coefficient. In Eq. 4, however, most of the computations are unnecessary, as what we need are only the diagonal elements. Thus, we reduce the computational overhead by solving it in a pixel-wise manner. Eq. 6 can then be solved by minimizing the construction dissimilarly with a L_0 sparse regularizer as:

$$\min_{m_p} \sum_i (||x_i - y_p^i \times m_p^i||_2^2 + \lambda||m_p^i||_0), \tag{12}$$

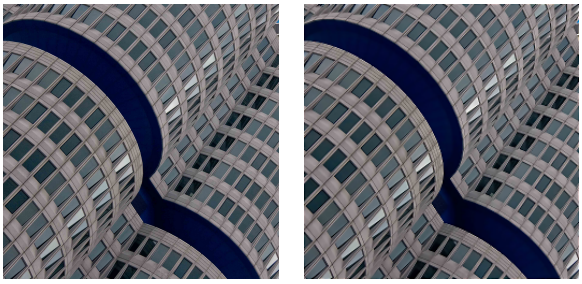


(a) Fixed output resolution.



(b) Fixed input resolution.

Fig. 4: Average computation time w.r.t. different numbers of input (a) and output (b) pixels.



(a) Original (b) [λ = 1.5e-2, σ = 4e-4]

(c) [λ = 1.5e-2, σ = 2e-2] (d) [λ = 4e-1, σ = 2e-2]

Fig. 5: Results generating by different parameter values. (Best viewed at full size on a high-resolution display.)

where i refers to the i^{th} column of m_p and the i^{th} row of y_p (as in Fig. 3), and x_i is a single pixel in downscaled image x. This pixel-wise computation allows the L_0 constraint to be applied to local patches, and thus to control the number of non-zero values within each patch in matrix m. To obtain an optimal solution for Eq. 12, we apply linear regression with a L_0-regularized term to approximate m_p, and the greedy matching pursuit (MP) algorithm [13] to compute the optimal sparse coefficients. Although MP is a local greedy algorithm and the solution is suboptimal, it has been shown to converge rapidly, in just 6 to 7 iterations [24]. The MP process is summarized in Algorithm 1. Parameter ξ is a predefined integer to constrain

(a) Original (b) Content-adaptive [10]

(c) Perceptual based [16] (d) Ours

Fig. 6: While existing methods suffer from inconsistent edges (b) and serious color splitting (c), the proposed method can preserve repeated patterns well (d).

the sparse elements of m_p^i, and τ is the allowed maximum residual to converge. m_p is first initialed based on gaussian distribution, and normalized to ensure that the sum of m_p^i is equal to 1. In our method, ξ has a fixed value of 0.6μ, and τ is set to 0.02 to balance between efficiency and performance.

V. Experiments

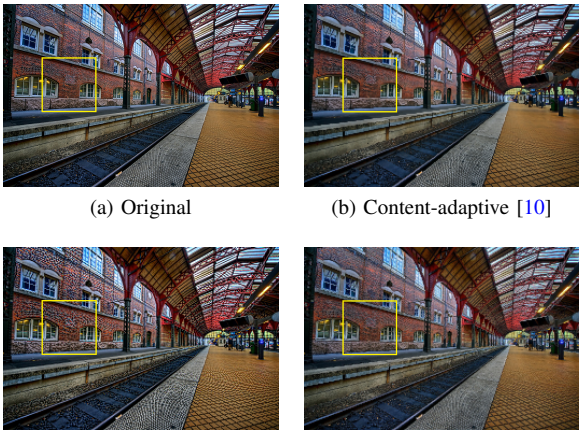
In this section, we first study the computation time of the proposed method. We then qualitatively evaluate the visual quality and examine the edge preservation of the results from the proposed method. In our evaluations, we compare the proposed method with two baseline methods, subsampling and bicubic, and two state-of-the-art methods, content-adaptive method [10] and perceptual based method [16].

A. Implementation Details

We have implemented the proposed method in Matlab, and tested it on a PC with an i7 3.2GHz CPU and 8GB RAM. It takes about 2 seconds to process a 500×400 image. We have further compared the computation time on 50 random images from the Urban100 dataset [9] and BSDS500 dataset [1]. The results are shown in Fig. 4. We fix the input/output size to find out the relationship between image size and computation time. Fig. 4a shows the result of fixing the output size to 80×60 while varying the input resolution, while Fig. 4b shows the result of fixing the input size to 640×480 while varying the output resolution. Results show that the computation time of the proposed method is independent of the downscaling factor, and has a $O(n)$ complexity to the number of input pixels. The proposed method has a lower computation time than the content-adaptive method [10], which computation time depends heavily on both the input size and downscaling factor. Comparing to the perceptual based method [16], the proposed method has a slightly higher computation time when the input image is not very large, but the difference becomes larger as the number of input pixels increases. However, for

(a) Original (b) Subsampling (c) Content-adaptive [10] (d) Perceptual based [16] (e) Ours

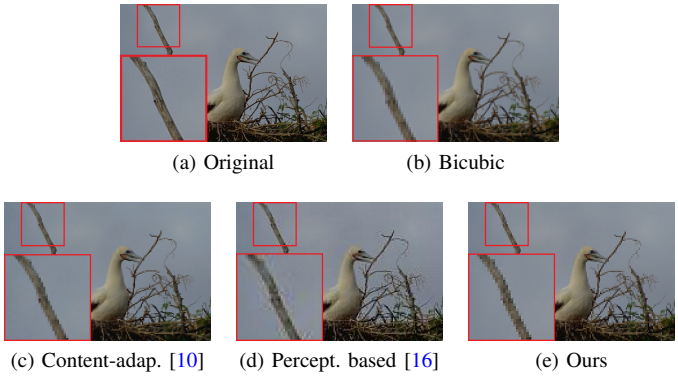
Fig. 7: City skyline examples. The proposed method is able to preserve the original visual perception with minimum artifacts such as color splitting and jagged edges.



(a) Original (b) Content-adaptive [10]

(c) Perceptual based [16] (d) Ours

Fig. 8: The proposed method preserves fine details on the brick wall (the yellow box) with sufficient perceptual contrast, while avoiding exaggerated textures shown in (c).



(a) Original (b) Bicubic

(c) Content-adap. [10] (d) Percept. based [16] (e) Ours

Fig. 9: Downscaling compressed images can be difficult. The image in (a) contains aliasing or ringing artifacts. The perceptual based method [16] in (c) amplifies the irrelevant compression artifacts.

the commonly used resolutions, the difference between these two methods is less than 1 second.

In Fig. 5, we show effects of different λ and σ values in an example. A higher λ value penalizes the intensity term in Eq. 12 and puts more emphasis on the sparse constraint of m_p, causing downsampling image y_l to be smoother as y_l is constructed from only a few pixels in y. Parameter σ mainly controls the degree of gradient magnitude that determines the salient edges. For all remaining experiments, we set $\lambda = 1.5e-2$, $\sigma = 2e-2$, and $\beta_{max} = 2$ (in Eq. 9) for better balancing between intensity and sparsity.

Note that all the experiments in this paper are based on a subsampling factor of 4, except Fig. 10. However, the proposed method can support both integer and non-integer downscaling factors. As the output size is controlled by the patch size and the patch step (see Fig. 3), non-integer downscaling can be achieved simply by allowing patches to overlap each other.

B. Qualitative Evaluations

In this section, we compare the proposed method with the content-adaptive [10] and the perceptual-based [16] methods, qualitatively through visual comparison and a user study. Tto evaluate the downscaling performance on handling luminance discontinuity of non-stochastic structures, we have selected images with structured regions from Urban100 [9], BSDS500 [1], and the Internet for the evaluations.

Visual Comparison: We have compared the proposed method with the existing methods on some challenging images. Fig. 6 and 7 show two images with structured details. We observe that the perceptual based method [16] produces color splitting artifacts around structured regions (e.g., new colors around the fence in Fig. 6 and the car grilles in first row of Fig. 1), as its optimization is performed on each color channel independently. The content-adaptive method [10] does

Fig. 10: The proposed downscaling method can perceptually preserve local details, even at different downscaling factors.

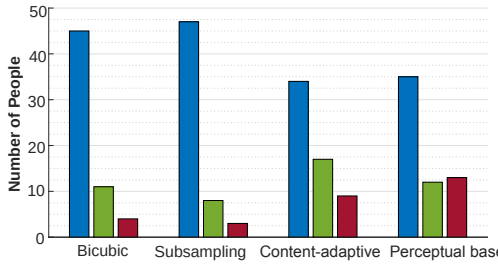


Fig. 11: User study results. Blue bars indicate the number of times that the proposed method is selected, and green bars indicate that of the competitors. Red bars indicate no preference.

not have from this problem, but it suffers from strong jagged edges (see Fig. 6) and may not preserve structured lines well (see Fig. 7). In contrast, the downscaled images from the proposed method have significantly fewer artifacts and are able to capture fine details well. Fig. 8 shows an example with fine repeated patterns. We can see that the perceptual based method [16] exaggerates the textured regions, while the content-adaptive method [10] fails to well preserve fine details resulting in over-smoothing. Fig. 9 shows a compressed image. We can see that the perceptual based method [16] exaggerates the compression noise due to the spatial-independent nature of the SSIM metric. Fig. 10 shows the results of the proposed method using different downscaling factors. We can see that the proposed method is robust to the the downscaling factor. In conclusion, the proposed method produces crispy results, but without the problems of the content-adaptive [10] and the perceptual based [16] methods.

User Study: We have conducted a user study to compare the output quality of the proposed method with the existing methods. The user study involves 13 natural images. For the sake of diversity, the test images are selected from three sources. We first randomly select 5 and 4 images from the Urban 100 and BSDS500 datasets, respectively. We then manually select 4 images from the Internet with totally different styles (e.g., painting) from the images already selected. Each participant requires an average of 20 minutes to finish this 13-image user study. Fig. 12 shows some of the test images. We have invited 60 participants for this study, and most of them were graduate students from different disciplines.

We have followed the evaluation setting in [10]. The user study is conducted in a pairwise comparison manner. We first show the original image and two downscaled images, one by our method and the other by one of the competitors. In each test, the participant may either select a preferred one from the two downscaled images or indicate no preference. All test images are displayed in native resolution, and zoom-in/zoom-out operations are not allowed. No time limit is set for each study. Other than these rules, the participants can view these images in their own preferred way, e.g., moving closer to or further away from the screen.

Fig. 11 shows the results of this user study. The proposed method (blue bars) achieves the best result in all pairwise comparisons. The content-adaptive method [10] achieves the closest performance to the proposed method, as it produces fewer artifacts than the perceptual based method [16] and subsampling, while preserving more details than bicubic. Feedbacks from the participants indicate that they concern about the crispness, jagged edges, and exaggerated textures found in some of the downscaled images. Although subsampling and the perceptual based method [16] produce more crispy downscaled images, their artifacts are noticeable. As a result, the proposed method achieves a significantly better performance over the state-of-the-art methods.

C. Contributions of the Priors

Here, we would like to evaluate the contributions of the priors. Since the proposed downsampling prior is essential to obtaining the downscaled images, we cannot remove it from the optimiation. Hence, Fig. 13 shows the results with and without the gradient-ratio prior only. We can see that without the gradient-ratio constraint, the downscaled image lacks fine details, while the proposed method preserves the details well.

D. Evaluation on Edge Preservation

Due to the lack of ground truth, we are unable to perform a quantitative evaluation on the proposed downscaling method. Hence, we evaluate it on edge preservation, as edges are one of the main visual perception indicators. This experiment can be considered as an extension of Fig. 2. We compare the proposed method with the state-of-the-arts on 100 BSDS [1] images with 3 downsampling factors (2, 4 and 8). The experimental settings are the same as those used in the edge detection benchmark BSDS [1]. Specifically, the edges of the resulting images are detected by Structured Edges [4]. The edge maps of different methods are upsampled and fed to the benchmark [1] to measure how well they match with the ground truth (i.e.,

(a) Original (b) Subsampling (c) Bicubic (d) Content-adap. [10] (e) Percept. based [16] (f) Ours

Fig. 12: Example images used in the user study.

salient edges labeled by human). Fig. 15 shows that the proposed method is able to preserve more edges than the state-of-the-art methods.

E. Limitations

Despite the advantages, the proposed method cannot preserve edges under all circumstances, especially for very fine edges, as shown in Fig. 16. These fine edges have smaller contrast and thus lower energies. Hence, they may be removed in the optimization process. As shown in Fig. 17, the repeated patterns with fine edges are removed. This is actually a common problem of all downscaling methods.

VI. Conclusion

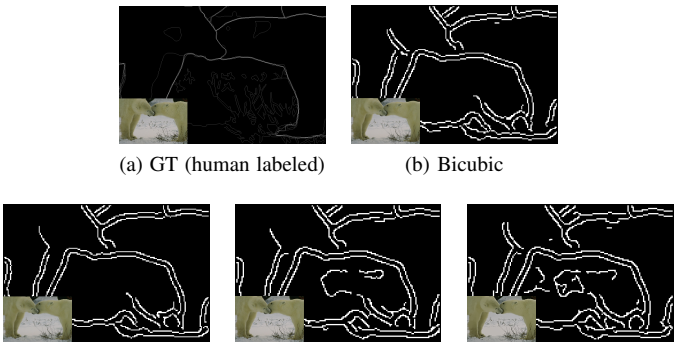
In this paper, we have proposed the L_0-regularized optimization framework for image downscaling. The framework is composed of two priors, gradient-ratio prior and reconstruction prior, which aim to preserve salient edges while maintaining the original structures. The proposed framework is able to eliminate downscaling artifacts such as aliasing, color splitting, and jagged edges. Finally, we show that the proposed image downscaling framework performs better than existing methods qualitatively as well as quantitatively.

References

[1] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. *IEEE TPAMI*, 33(5):898–916, 2011.

(a) Original (b) Ours without gradient-ratio prior (c) Ours

Fig. 13: Contribution of the proposed gradient-ratio prior. It preserves most of the fine details in the downscaled image.



(a) GT (human labeled) (b) Bicubic

(c) Content-adap. [10] (d) Percep. based [16] (e) Ours

Fig. 14: Edge preservation evaluation. Salient edges are extracted using Structured Edges [4]. We then evaluate the resulting edge maps with the BSDS benchmark [1] on the accuracy of preserving salient edges.

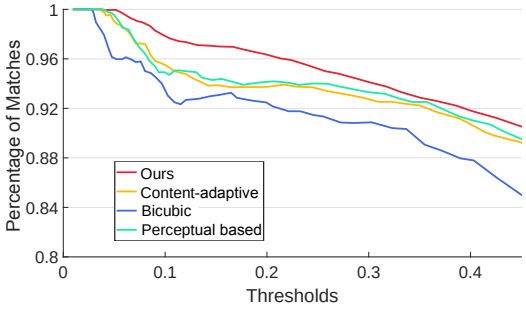
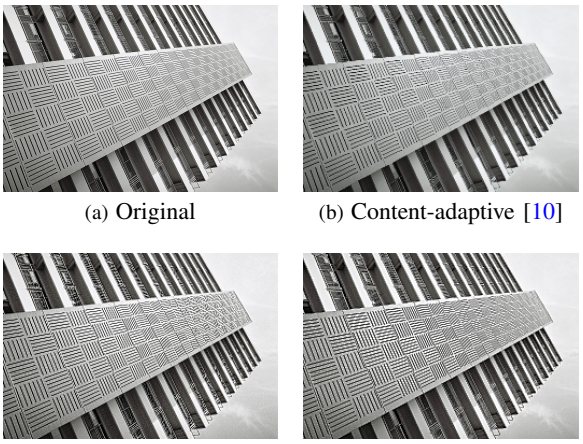


Fig. 15: Result of the BSDS benchmark [1]. The proposed L_0 algorithm preserves more edges than the state-of-the-arts.



(a) Original (b) Content-adaptive [10]

(c) Perceptual based [16] (d) Ours

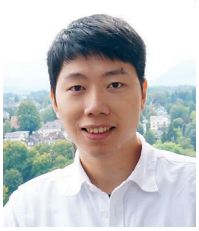
Fig. 16: Limitation of the proposed method. It fails to preserve fine and non-salient edges.

Fig. 17: Fine repeated patterns are very difficult to preserve.

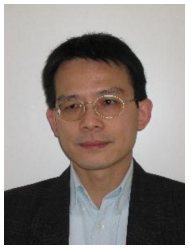
[2] S. Avidan and A. Shamir. Seam carving for content-aware image resizing. *ACM TOG*, 26(3), 2007.
[3] D. Burr, M. Morrone, and D. Spinelli. Evidence for edge and bar detectors in human vision. *Vision Research*, 29, 1989.
[4] P. Dollár and C. L. Zitnick. Fast edge detection using structured forests. *IEEE transactions on pattern analysis and machine intelligence*, 37(8):1558–1570, 2015.
[5] Y. Fang, Z. Chen, W. Lin, and C. Lin. Saliency detection in the compressed domain for adaptive image retargeting. *IEEE TIP*, 21(9):3888–3901, 2012.
[6] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski. Edge-preserving decompositions for multi-scale tone and detail manipulation. In *ACM TOG*, volume 27, page 67. ACM, 2008.
[7] E. S. Gastal and M. M. Oliveira. Domain transform for edge-aware image and video processing. In *ACM TOG*, volume 30, page 69. ACM, 2011.
[8] K. He, J. Sun, and X. Tang. Guided image filtering. *IEEE TPAMI*, 35(6):1397–1409, 2013.
[9] J. Huang, A. Singh, and N. Ahuja. Single image super-resolution from transformed self-exemplars. In *Proc. CVPR*, pages 5197–5206, 2015.
[10] J. Kopf, A. Shamir, and P. Peers. Content-adaptive image downscaling. *ACM TOG*, 32(6), 2013.
[11] A. Liu, W. Lin, and M. Narwaria. Image quality assessment based on gradient similarity. *IEEE TIP*, 21(4):1500–1512, 2012.
[12] R. Lukac. *Perceptual Digital Imaging: Methods and Applications*. CRC

Press, 2012.
[13] S. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Trans. on Signal Processing*, 41(12):3397–3415, 1993.
[14] D. Mitchell and A. Netravali. Reconstruction filters in computer-graphics. In *Proc. ACM SIGGRAPH*, pages 221–228, 1988.
[15] D. Nehab and H. Hoppe. Generalized sampling in computer graphics. Technical report, MSR-TR-2011-16, 2011.
[16] A. Öztireli and M. Gross. Perceptually based downscaling of images. *ACM TOG*, 34(4), 2015.
[17] M. Rubinstein, A. Shamir, and S. Avidan. Multi-operator media retargeting. 28(3):23, 2009.
[18] R. Samadani, S. Lim, and D. Tretter. Representative image thumbnails for good browsing. In *Proc. ICIP*, volume 2, 2007.
[19] R. Shapley and D. Tolhurst. Edge detectors in human vision. *The Journal of Physiology*, 229, 1973.
[20] B. Suh, H. Ling, B. Bederson, and D. Jacobs. Automatic thumbnail cropping and its effectiveness. In *Proc. UIST*, pages 95–104, 2003.
[21] P. Thévenaz, T. Blu, and M. Unser. Interpolation revisited. *IEEE Trans. on Medical Imaging*, 19(7):739–758, 2000.
[22] T. Troscianko, C. Benton, P. Lovell, D. Tolhurst, and Z. Pizlo. Camouflage and visual perception. *Philosophical Trans. of the Royal Society B: Biological Sciences*, 364(1516):449–461, 2009.
[23] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE TIP*, 13(4):600–612, 2004.
[24] J. Xie, R. Feris, S. Yu, and M. Sun. Joint super resolution and denoising from a single depth image. *IEEE TMM*, 17(9):1525–1537, 2015.
[25] L. Xu, C. Lu, Y. Xu, and J. Jia. Image smoothing via L0 gradient minimization. *ACM TOG*, 30(6), 2011.

Junjie Liu is a Research Assistant at South China University of Technology. He obtained his B.Sc. degree from South China Agricultural University, and the M.Sc. degree from City University of Hong Kong. His research interests include computer vision and image processing.



Shengfeng He is an Associate Professor in the School of Computer Science and Engineering, South China University of Technology. He was a Research Fellow at City University of Hong Kong. He obtained his B.Sc. degree and M.Sc. degree from Macau University of Science and Technology, and the Ph.D degree from City University of Hong Kong. His research interests include computer vision, image processing, computer graphics, and deep learning.



Rynson W.H. Lau received his Ph.D. degree from University of Cambridge. He was on the faculty of Durham University and is now with City University of Hong Kong.

Rynson serves on the Editorial Boards of Computer Graphics Forum, and Computer Animation and Virtual Worlds. He has served as the Guest Editor of a number of journal special issues, including ACM Trans. on Internet Technology, IEEE Trans. on Multimedia, IEEE Trans. on Visualization and Computer Graphics, and IEEE Computer Graphics & Applications. He has also served in the committee of a number of conferences, including Program Co-chair of ACM VRST 2004, ACM MTDL 2009, IEEE U-Media 2010, and Conference Co-chair of CASA 2005, ACM VRST 2005, ACM MDI 2009, ACM VRST 2014. Rynson's research interests include computer graphics, image processing and computer vision.