

Deformation Aware Image Compression

Tamar Rott Shaham and Tomer Michaeli
Technion—Israel Institute of Technology, Haifa, Israel
{stamarot@campus, tomer.m@ee}.technion.ac.il

Abstract

Lossy compression algorithms aim to compactly encode images in a way which enables to restore them with minimal error. We show that a key limitation of existing algorithms is that they rely on error measures that are extremely sensitive to geometric deformations (e.g. SSD, SSIM). These force the encoder to invest many bits in describing the exact geometry of every fine detail in the image, which is obviously wasteful, because the human visual system is indifferent to small local translations. Motivated by this observation, we propose a deformation-insensitive error measure that can be easily incorporated into any existing compression scheme. As we show, optimal compression under our criterion involves slightly deforming the input image such that it becomes more “compressible”. Surprisingly, while these small deformations are barely noticeable, they enable the CODEC to preserve details that are otherwise completely lost. Our technique uses the CODEC as a “black box”, thus allowing simple integration with arbitrary compression methods. Extensive experiments, including user studies, confirm that our approach significantly improves the visual quality of many CODECs. These include JPEG, JPEG 2000, WebP, BPG, and a recent deep-net method.

1. Introduction

The last decades have seen an exponential rise in the popularity of mobile devices equipped with high-resolution cameras. To accommodate the numerous amounts of pictures captured by those devices on a daily basis, there is a crucial need for high quality compression algorithms. Indeed, while 20 megapixel images are becoming common (requiring 60 megabytes to store uncompressed), transmission and storage is often limited to less than 1 megabyte per image. At such high ratios, commonly used compression methods tend to discard important information from the image and produce visually unpleasing results.

In this paper, we propose a generic approach for boosting the visual quality of any image compression method, by introducing deformations to the input image (see Fig. 1). Our

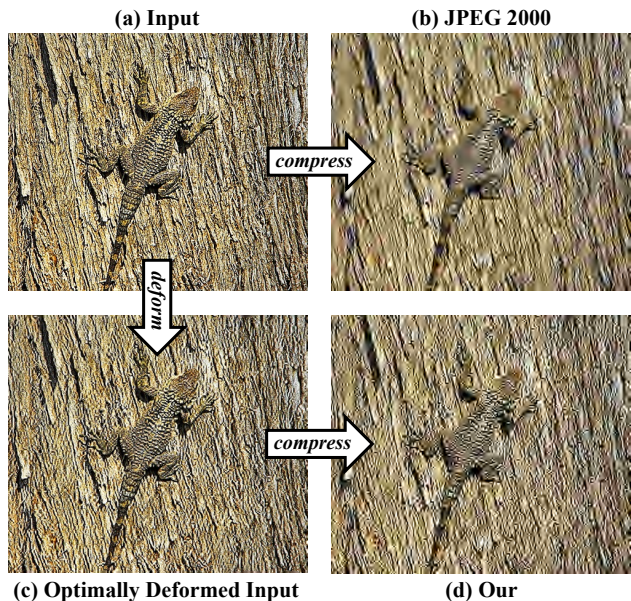


Figure 1. **Deformation aware image compression.** Our algorithm seeks to minimize a deformation-insensitive error measure. This boils down to determining how to best deform the input image (a) so as to make it more compressible (c). By doing so, we trade a little geometric integrity with a significant gain in terms of preservation of visual information (d) compared to regular compression (b). Note that images (b) and (d) were obtained by compressing images (a) and (c), respectively, with the *same* JPEG 2000 CODEC using the *same* compression ratio of 150:1.

algorithm uses the CODEC as a “black box” and is thus extremely simple to incorporate into arbitrary methods. Yet, it has a pronounced effect: At the same bit rate, we are able to achieve significantly better visual results.

Lossy compression schemes attempt to compactly encode images in a way which allows to restore them with minimal error. Over the years, most efforts to improve compression methods focused on seeking better image models. Examples include sparsity of image blocks in the DCT domain [35], recurrence of patches across different scales of the image [3], sparsity in the wavelet domain [28], and smoothness (as exploited e.g. by PDE based

approaches [8, 25]). However, the impact of new image models seems to be slowing down. Indeed, newer and more sophisticated priors now outperform their predecessors by small margins and only at high compression ratios [1, 4].

Here, we take a different route. Rather than focusing on the image prior, we focus on the error criterion. Specifically, most compression methods seek to minimize some per-pixel distance (typically ℓ_2) between the input image and the decoded image. Several attempts were also made to incorporate the SSIM index [39] as a fidelity criterion, leading to only modest improvement in visual quality [21, 10]. We claim that the main limitation of most existing distance measures (including perceptual ones) is that they are very sensitive to slight misalignment of shapes and objects in the two images. Therefore, excelling under those criteria requires encoding the precise geometry of every fine detail in the image. But this is clearly wasteful, as the human visual system is not distracted by small geometric deformations, as long as the semantics of the scene is preserved.

Motivated by this insight, in this paper we propose a new error measure, which is insensitive to small smooth deformations. Our measure has two key advantages over other criteria: (i) it is very simple to incorporate into any compression method, and (ii) in the context of compression, it better correlates with human perception (as we confirm by user studies), and thus leads to a significant improvement in terms of detail preservation.

As we show, optimal compression under our criterion boils down to determining how to best deform the input image such that it becomes more “compressible”. This is illustrated in Figs. 1 and 2 for the JPEG 2000 [28] and the Global Thresholding [29] compression methods. As can be seen, by introducing very minor deformations, we are able to make the compression scheme preserve delicate features that are otherwise completely lost. Note that this effect is achieved without increasing the bit rate. In other words, rather than discarding textures and small objects to meet the bit budget, we geometrically modify them such that they can be better encoded with the same number of bits.

The surprising success of our approach can be attributed to an interesting phenomenon recently observed in [23]. That is, by introducing small geometric deformations, it is usually possible to significantly increase the likelihood of any natural image under any given prior. The implication of this effect on compression is striking. For example, compression algorithms that exploit sparsity in the wavelet domain (e.g. JPEG 2000), discard the small wavelet coefficients of the image. At high compression ratios, this causes fine details to fade, as demonstrated in Fig. 3 for the Sub-band Thresholding compression method [29]. However, as can be seen in Fig. 3(b),(d), it takes only a small deformation to make the wavelet transform of the image significantly sparser. Thus, by slightly sacrificing geometric in-

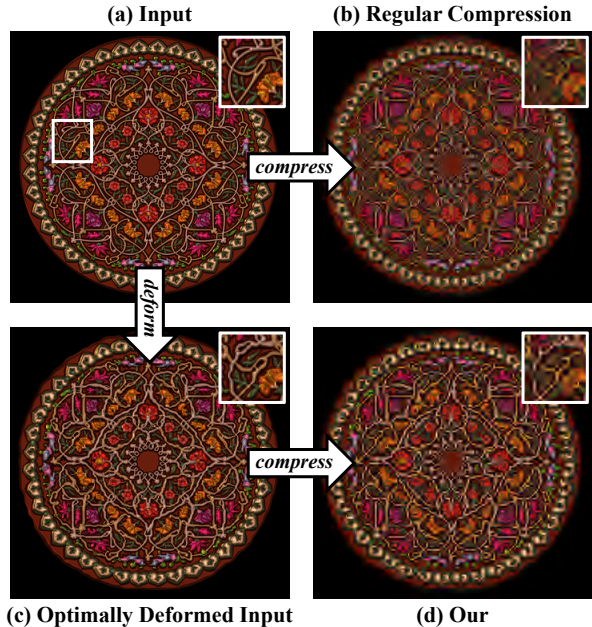


Figure 2. **The effect of deformation.** When compressing the input image (a) using the Global Thresholding method [18] at a ratio of 40:1, delicate structures are completely lost (b). Interestingly, by making those structures a bit wiggly (c) they can be significantly better preserved during compression (d).

tegrity, we substantially improve the ability of the compression algorithm to preserve details, as seen in Fig. 3(f).

2. Related Work

The most popular error measure in image compression is the squared ℓ_2 distance. This criterion is mathematically convenient, being convex and differentiable, but is unarguably not well correlated with subjective human perception of image quality [38, 14, 42].

There exist many fidelity criteria that are better correlated with human perception. A few examples are SSIM [39], MS-SSIM [40], CW-SSIM [24], IFC [27], VIF [26], and FSIM [41]. Recently, several perceptual loss functions have been proposed, which measures the similarity between deep feature maps (mostly of the VGG net). These measures were shown to lead to pleasing visual results in a variety of low-level vision tasks, including super-resolution [13, 11] and style transfer [16]. Perceptual losses were also incorporated with generative adversarial networks (GANs), allowing to achieve high-quality super-resolution [13] and compression [22] results.

While these criteria better match human perception, their majority lack a crucial property for perceptual compression: *deformation invariance*. Namely, they do not tolerate small misalignment of objects, and thus necessitate the encoder to invest many bits in encoding the fine geometry of every feature in the image. Indeed, several attempts to incorpo-

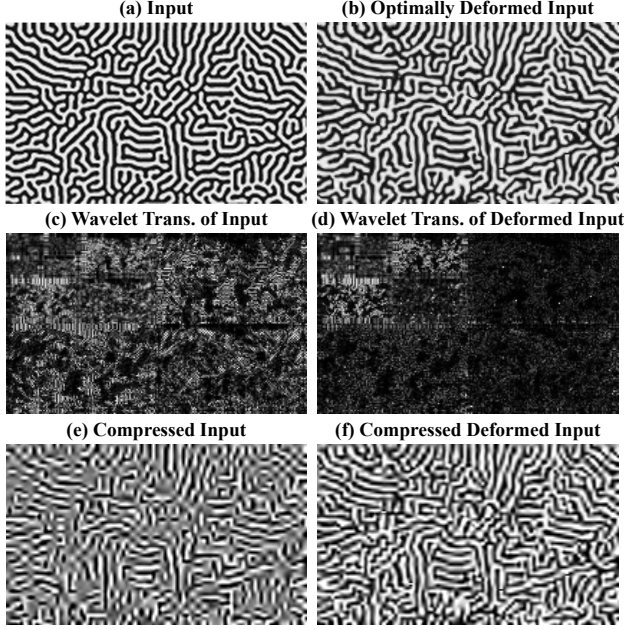


Figure 3. **Deformation aware compression via Subband Thresholding.** The input image (a) contains many strong curved edges, so that its wavelet transform (c) is not very sparse. This causes Subband Thresholding compression [29] at a ratio of 25:1, to produce a very blurry result (e). However, by introducing a minor geometric deformation (b), we are able to make the wavelet transform of the image much sparser (d). This allows the compression algorithm to preserve most of the structures in the image, at the same bit budget (f). This principle applies to arbitrary image priors, and not only those based on wavelet sparsity.

rate the SSIM criterion into JPEG and JPEG 2000 [21, 10] and into video coding [36, 20, 37], led to rather modest improvements in visual quality.

Another drawback of existing error measures is that they are difficult to incorporate into arbitrary compression schemes. Namely, as opposed to the ℓ_2 distance, rate-distortion optimization under those measures cannot be done analytically and thus requires various approximations. Thus, even when solutions exist [21, 10, 20], they are quite specific and cannot be easily extended to other compression standards. Furthermore, from an end-user viewpoint, those solutions require a specialized CODEC. Our solution, on the other hand, can work with any existing CODEC. Specifically, we use the CODEC as a “black box”, to generate a preprocessed (deformed) image. This image can then be compressed and decompressed using the original CODEC, without any modification.

Our method is related to a recent line of work on using deformations for idealizing images [5, 33, 23]. These papers introduced the idea of deforming images as a means for making them better comply with some prior model. Here, we harness this idea for improving image compres-

sion. Namely, we propose an error criterion which measures similarity up to small deformations. Thus, compression under our criterion, reduces to determining how the input image should be deformed so that it is more compressible.

Note that the idea of measuring image similarity up to deformation has been proposed in the context of image recognition [19] and face recognition [34]. However, this approach has never been exploited for image compression.

3. Deformation Aware Compression

Modern compression schemes involve a procedure known as rate-distortion optimization. Namely, during compression, the algorithm adaptively selects where to invest more bits so as to minimize the distortion between the input image y and its compressed version x , while conforming to a total bit rate constraint of ε bits per pixel. This can be formulated as the optimization problem

$$\min_x d(x, y) \quad \text{s.t.} \quad R(x) \leq \varepsilon, \quad (1)$$

where $d(\cdot, \cdot)$ is some distortion measure that quantifies the dissimilarity between x and y , and $R(x)$ is the rate required to encode x .

The most popular distortion measure is the sum of squared differences (SSD), *i.e.* the square ℓ_2 error norm

$$d_{\text{SSD}}(x, y) = \|x - y\|^2. \quad (2)$$

The SSD is a per-pixel criterion, and is therefore extremely sensitive to slight misalignment or deformation of objects. For that reason, when the bit budget ε is low (*i.e.* high compression ratio), the compression process completely removes or blurs out certain structures in the image.

As human observers are indifferent to slight local translations, here we propose a deformation insensitive version of the SSD measure. We consider two images x and y to be similar, if there exists a smooth deformation \mathcal{T} such that x and $\mathcal{T}\{y\}$ are similar. More concretely, we define the deformation aware SSD (DASSD) between x and y as

$$d_{\text{DASSD}}(x, y) = \min_{\mathcal{T}} \|x - \mathcal{T}\{y\}\|^2 + \lambda \psi(\mathcal{T}), \quad (3)$$

where the term $\psi(\mathcal{T})$ penalizes for non-smooth deformations. Note that computing the DASSD requires solving an optical-flow problem [9] to determine how to best warp y onto x . Once the optimal deformation is determined, the DASSD is defined as the SSD between x and the warped y , plus a term that quantifies the roughness of the flow field. The parameter λ controls the tradeoff between the two terms. Therefore, the DASSD is large if the best warped y is not similar to x , or if the deformation required to make y similar to x is not smooth (or both).

To allow for complex deformations, we use a nonparametric flow field (u, v) , namely

$$\mathcal{T}\{y\}(\xi, \eta) = y(\xi + u(\xi, \eta), \eta + v(\xi, \eta)). \quad (4)$$



	Similarity scores (higher = similar)						Dissimilarity scores (lower = similar)			
	SSIM	MS-SSIM	CW-SSIM	IFC	VIF	FSIM	VGG _{2,2}	VGG _{5,4}	SSD	DASSD (our)
JPEG 2000	0.696	0.837	0.999	1.216	0.124	0.857	0.173	2.023	545	328
Our	0.685	0.810	0.996	1.014	0.104	0.848	0.252	2.347	639	284

Figure 4. **Comparison between error criteria.** We measured the similarity between the input image (a) and the compressed images (b) and (c), according to various fidelity criteria. We indicate in bold which compression result has been ranked more similar: our deformation aware compression (c), or the regular JPEG 2000 compression (b) (both compressed at a ratio of 200:1). As can be seen, all error criteria other than DASSD rank our result as less similar to the input image. This is despite the fact that it preserves much more visual information.

We define the penalty $\psi(\mathcal{T})$ to be a weighted Horn and Schunk regularizer [9],

$$\psi(\mathcal{T}) = \iint w(\xi, \eta) (\|\nabla u(\xi, \eta)\|^2 + \|\nabla v(\xi, \eta)\|^2) d\xi d\eta, \quad (5)$$

where $\nabla = (\frac{\partial}{\partial \xi}, \frac{\partial}{\partial \eta})$ and $w(\xi, \eta)$ is a weight map that puts higher penalty on salient regions (see Sec. 4).

Figure 4 illustrates the advantage of our DASSD criterion over other similarity criteria. In this example, our method clearly preserves more details than the original JPEG 2000 compression. This is captured by our DASSD measure, which ranks our compressed image as more similar to the input image than the regular JPEG 2000 result (here we used a constant regularization map $w(\xi, \eta) = 1$). However, the other similarity measures are very sensitive to misalignment of objects, and thus all of them rank our result as less similar to the input image. Please see a comparison of all similarity measures on all the images in this paper in the Supplementary Material.

Our measure is obviously the least sensitive to smooth deformations. But why should deformation invariance improve compression? Lossy image compression schemes are usually not translation invariant. That is, compressing a shifted version of an image, gives an entirely different result than shifting the compressed image. This is demonstrated in Fig. 5 for the JPEG 2000 standard. While the input image and its shifted version look perfectly identical to a human observer, their compressed versions look very different. In one of them the small square in the middle is preserved, and in the other it is not. As opposed to SSD, our deformation aware criterion prefers the result in which the small square is preserved: The DASSD between (a) and (d) is 3% lower

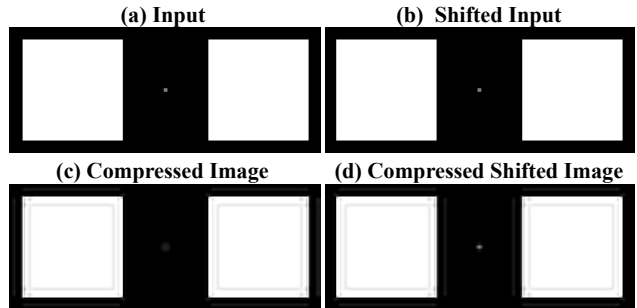


Figure 5. **The effect of global translation.** Compressing the image (a) using JPEG 2000 at a ratio of 75:1, causes the small square in the middle to disappear (c). However, by shifting the image only two pixels to the left (b), compression at the same ratio keeps the small square intact (d).

than the DASSD between (a) and (c), while the SSD between (a) and (d) is 16 times larger than the SSD between (a) and (c). This intuition can be extended to local translations. For example, to preserve the pattern of the Mandala in Fig. 2, it is necessary to make lines a bit wiggly. These small translations make the image more compressible, thus leading to better visual quality at the same bit rate.

4. Algorithm

Substituting d_{DASSD} of (3) into (1), we obtain the optimization problem

$$\min_{x, \mathcal{T}} \|\mathcal{T}\{y\} - x\|^2 + \lambda \psi(\mathcal{T}) \quad \text{s.t.} \quad R(x) \leq \varepsilon. \quad (6)$$

That is, we need to simultaneously determine a compressed image x (represented with no more than ε bits per pixel)

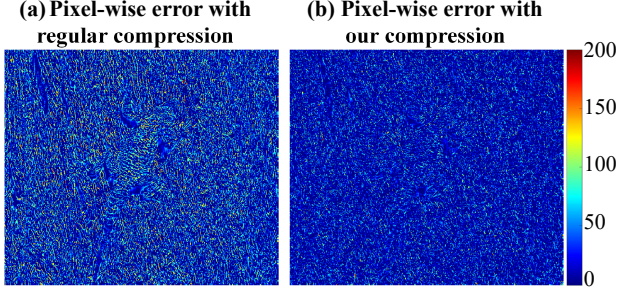


Figure 6. **Pixel-wise compression error.** Our algorithm deforms images so as to make them more compressible in the ℓ_2 sense. The effect is that the error obtained in compressing the deformed image (b) is significantly lower than the error obtained in compressing the original image (a). These results correspond to the Lizard images of Fig. 1(c) and Fig. 1(a), respectively.

and a geometric deformation \mathcal{T} , such that x is similar to the deformed image $\mathcal{T}\{y\}$ rather than to y itself. In other words, we seek how to deform the input image y , such that $\mathcal{T}\{y\}$ can be compressed with smaller SSD error under the same bit budget. This is illustrated in Fig. 6.

To solve problem (6) we alternate between minimizing the objective w.r.t. x while holding \mathcal{T} fixed and vice versa.

x -step: When \mathcal{T} is fixed, $\psi(\mathcal{T})$ can be discarded, so that (6) simplifies to

$$\min_x \|x - \mathcal{T}\{y\}\|^2 \quad \text{s.t.} \quad R(x) \leq \varepsilon. \quad (7)$$

This is a standard rate distortion problem, but for compressing the deformed image $\mathcal{T}\{y\}$ rather than the input image y .

\mathcal{T} -step: When x is fixed, the bit rate $R(x)$ is constant, and (6) reduces to

$$\min_{\mathcal{T}} \|x - \mathcal{T}\{y\}\|^2 + \lambda \psi(\mathcal{T}). \quad (8)$$

This is an optical flow problem [9] for determining how to best warp the input image y onto the compressed image x . Here we use the iteratively re-weighted least-squares (IRLS) algorithm proposed in [15].

Thus, as summarized in Alg. 1, our algorithm iterates between two simple steps: Compressing the current deformed input image $\mathcal{T}\{y\}$ to obtain x , and computing the optical flow between x and y to update \mathcal{T} . Note that the x -step can be done with any CODEC. This allows integrating our algorithm with arbitrary compression methods.

To prevent the algorithm from getting trapped in a bad local minimum, we start with a large bit budget ε , and gradually decrease it along the iterations until we reach the desired budget. This helps to avoid the following situation. Suppose that at some stage, the compression (x -step) removes some structure from the image, so that it appears in $\mathcal{T}\{y\}$ but not in x . In that case, the optical flow (\mathcal{T} -step)

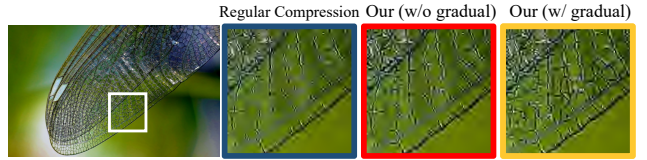
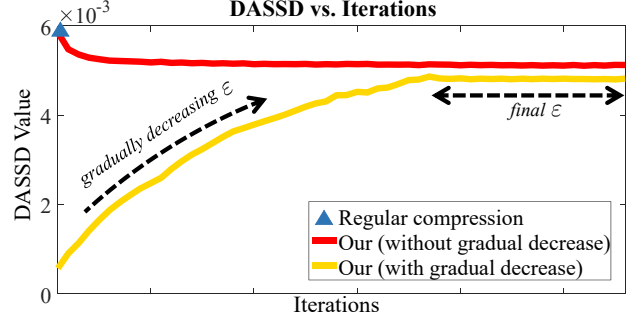


Figure 7. **The effect of gradual bit rate decrease.** Our algorithm iterates between compression and optical flow estimation. When the compression step is performed with the final desired bit rate (red box), the algorithm fails to restore all the fine structures on the Dragonfly’s wing, that are lost with regular compression (blue box). However, when starting with a large bit rate and progressively decreasing it, our algorithm manages to preserve most of the wing’s delicate textures (yellow box). This is correlated with the final DASSD value, which is 10% lower when using the gradual scheme (yellow line), compared to the direct scheme (red line). We used the JPEG 2000 standard with a compression ratio of 150:1.

```

Input: Image  $y$ , bit budget  $\varepsilon$ 
Output: Compressed image  $x$ 
Initialize  $\mathcal{T}$  to the identity mapping and  $\tilde{\varepsilon}$  to be large
Compute the local regularization weight map  $w$ 
while  $\tilde{\varepsilon} > \varepsilon$  do
     $x \leftarrow \text{Compress}(\mathcal{T}\{y\}, \tilde{\varepsilon})$  /* bit budget of  $\tilde{\varepsilon}$  */
     $\mathcal{T} \leftarrow \text{OpticalFlow}(y, x)$  /* with weight map  $w$  */
    decrease  $\tilde{\varepsilon}$ 
end

```

Algorithm 1: Deformation Aware Compression.

cannot determine how to best deform this structure so as to encourage the compression to preserve it in the next iteration. Our gradual process overcomes this issue by allowing the deformation to gradually adapt to the small structures before they disappear. Therefore, as we show in Fig. 7, this process leads to better detail preservation. Indeed, with the gradual scheme, the algorithm converges to a DASSD value which is 10% lower than without the gradual scheme. Note that the low DASSD values at early iterations are a result of using a high bit rate. As the bit rate decreases, the DASSD values increase (but reach a lower value at the final bit rate, than without the gradual process).

Constructing the regularization weight map To ensure good visual quality, we need to prevent extreme distortions

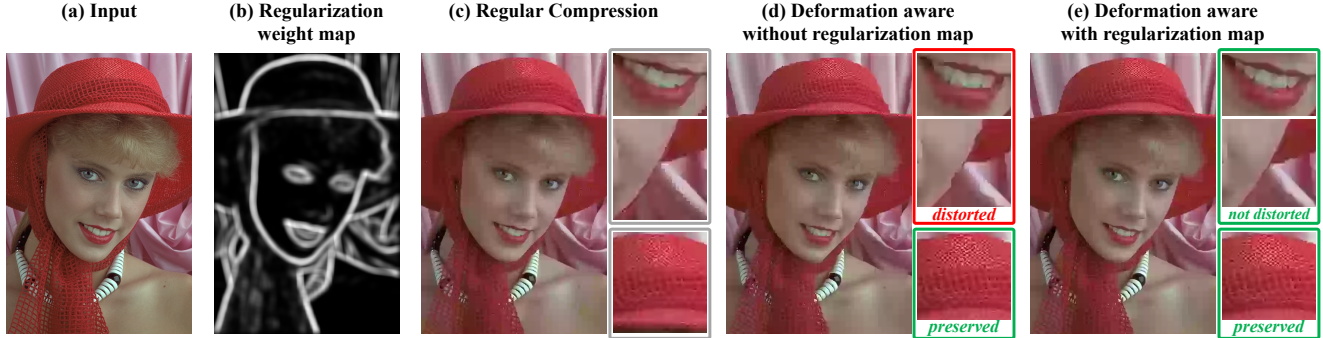


Figure 8. **The effect of the regularization weight map.** Applying our scheme (with WebP compression at a ratio of 250:1) on image (a) without adaptive regularization (constant $w(\xi, \eta)$), results in image (d) in which the hat’s texture is better preserved, compared to the regular WebP compression (c), but the woman’s lips are distorted unrealistically. Using the local regularization weight map (b), results in image (e) in which the boundaries of the lips are not distorted, yet the texture on the hat is still sharper than with the regular compression (c).

at regions which capture the observer’s attention. In particular, humans are very sensitive to the outline of objects. We therefore construct the weight map $w(\xi, \eta)$ in (5) as

$$w(\xi, \eta) = 1 + \alpha \cdot G(\xi, \eta) * E(\xi, \eta), \quad (9)$$

where E is an edge map obtained by applying the edge detector of [6] on the input image y , G is a Gaussian filter with $\sigma = 20$, ‘*’ denotes convolution, and α is a parameter that controls the strength of the varying regularization. As can be seen from Fig. 8, using local regularization is essential for avoiding distracting artifacts. With a global regularization (constant w), object boundaries are distorted unrealistically (*e.g.* the woman’s lips). By introducing a spatially varying regularization, lines and boundaries are not distorted, yet textures (*e.g.* the hat) are still allowed to deform and are thus better preserved during compression.

5. Experiments

We tested our approach with JPEG [35], JPEG 2000 [28], WebP [1], BPG [4], the deep-net based CODEC of [32], Subband Thresholding [29] and Global Thresholding [18], on images from the Berkeley segmentation dataset [17], the Kodak dataset [7] and the Web (please see many more results in the Supplementary Material). For the gradual process, we kept the compression ratio fixed for the first 10 iterations, then increased it every 5 iterations for the next 25 iterations, and then increased it every single iteration until reaching the desired rate. For JPEG 2000, we started at a compression ratio of 20:1 and increased it by steps of 5. For Subband Thresholding and Global Thresholding we started at a ratio of 5:1 and increased it by steps of 1. For the deep coding algorithm, we started with a bit rate of 0.75 bits per pixel (BPP), and then decreased it by steps of 0.125 BPP. In JPEG, BPG and WebP, the user specifies a quality parameter rather than the desired compression rate. Thus, for JPEG and WebP, we

started at a quality of 50 and decreased it by steps of 1. For BPG we started at a quality index of 30 (here a lower index corresponds to better quality) and increased it by steps of 1. In each optical flow step, we used the flow from the previous iteration as initialization. We found this leads to better convergence. All warped images were produced with bicubic interpolation (this induces negligible blur which does not affect the visual quality).

The local geometries preferred by different compression schemes are of different nature. In some cases, those preferred structures look quite unnatural to the human eye (*e.g.* blockiness effects in JPEG). To prevent our approach from generating unpleasant images, we tuned the parameter α of the regularization map (9) differently for different compression schemes. For JPEG we used $\alpha = 20$, for JPEG 2000 and Global and Subband thresholding we used $\alpha = 3$, and for WebP, BPG and Deep Coding we used $\alpha = 6$. For all compression methods we used $\lambda = 65$ (for color images with 8 bits per pixel per channel).

The running time of our algorithm is given by $T = K \times (T_{\text{CODEC}} + T_{\text{flow}})$, where K is the number of iterations and $T_{\text{CODEC}}, T_{\text{flow}}$ are the running times of the CODEC and optical flow, respectively. Typically, K is on the order of a few tens, $T_{\text{flow}} \approx 0.5$ sec and $T_{\text{CODEC}} \approx 0.15$ sec for a 1 megapixel image.

Figures 2 and 3 depict results produced by our algorithm with the Subband Thresholding [29] and the Global Thresholding [18] compression methods, respectively. These two simple approaches produce unpleasing visual results already at moderate compression ratios. However, by using our approach, we are able to improve their performance. This demonstrates that the choice of the error criterion is not less important than the choice of the image prior. Indeed, even simple models can lead to good visual results when used with a deformation indifferent error criterion.

Figures 1, 9 and 10 show several results produced by our algorithm with the JPEG and JPEG 2000 schemes. As can



Figure 9. **JPEG 2000** [28]. Compression of Houses at a ratio of 50:1.

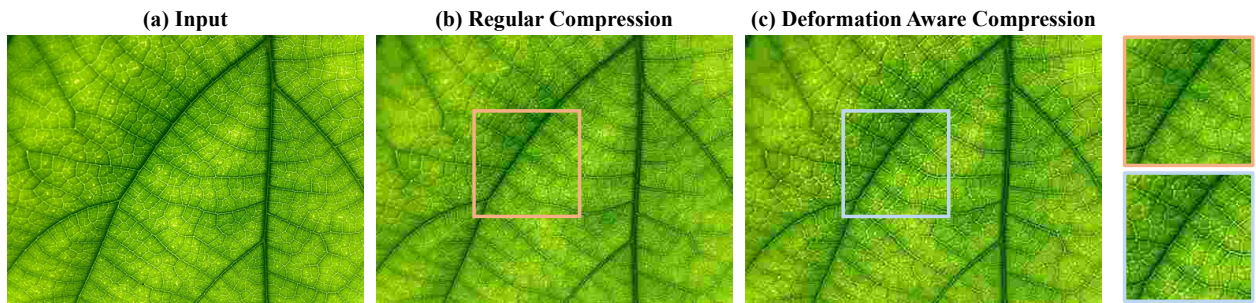


Figure 10. **JPEG** [35]. Compression of Leaf at a ratio of 50:1.

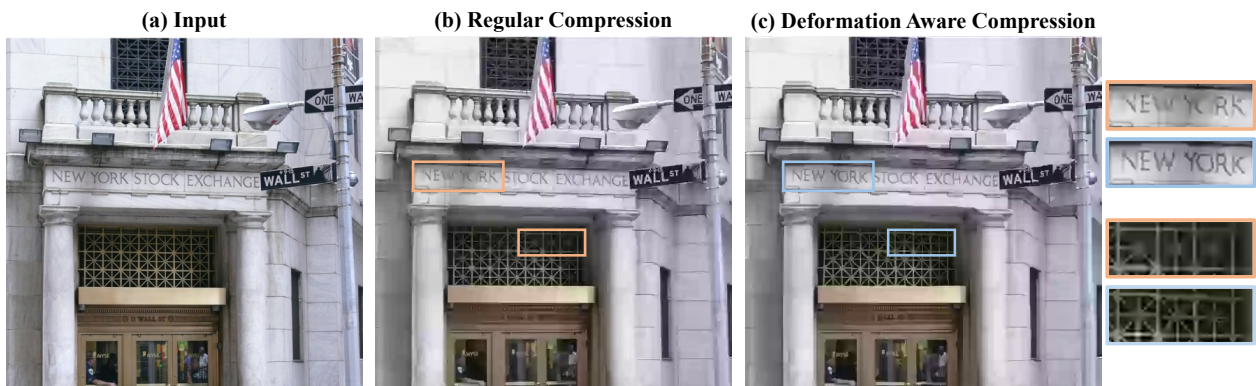


Figure 11. **WebP** [1]. Compression of the New York Stock Exchange building at a ratio of 110:1.

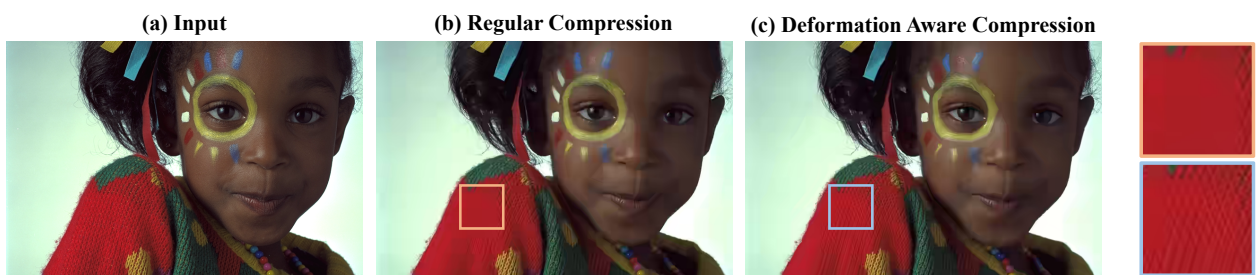


Figure 12. **BPG** [4]. Compression of Girl at a ratio of 220:1.

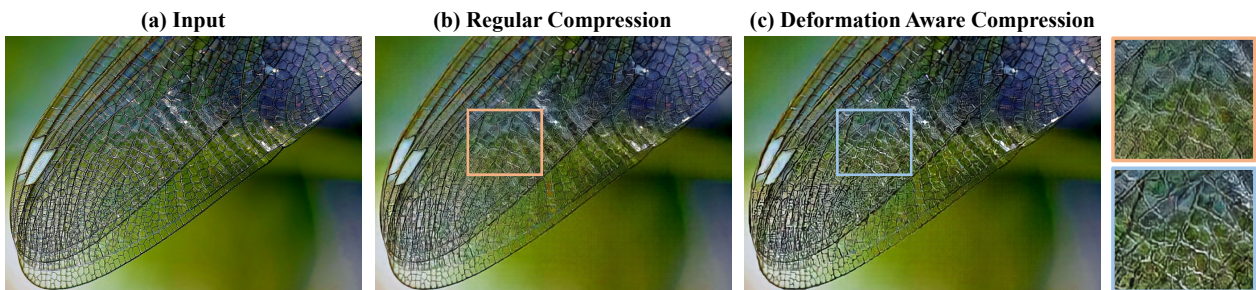


Figure 13. **Deep Coding** [32]. Compression of Dragonfly's wings at a ratio of 48:1.

be seen, our algorithm has a very pronounced visual effect: It manages to preserve a lot of the content that is completely lost in regular compression. In particular see the house’s roof and windows and the threads on the leaf.

Next, we applied our algorithm on the newer compression methods WebP and BPG. Generally, we found that the improvement for those methods is moderate and sometimes even unnoticeable. However, in some cases (e.g. Figs. 11 and 12) the contribution of our approach is extremely meaningful. Note how our algorithm restores fine details like the words ‘NEW YORK’ and the texture on the sweater, that are otherwise completely dissolved.

Recently, several neural net based lossy compression methods have been proposed [22, 32, 12, 2, 30, 31]. To test the effect of deformation awareness on this family of techniques, we experimented with the CODEC of [32]. As illustrated in Fig. 13, our approach significantly boosts the visual quality of this method. This suggests that our method is also of great relevance to the recent trend of deep-net based compression.

As demonstrated in Fig. 4, since our method introduces deformations, most error criteria tend to rank our results as less similar to the input image than the regular compression. Therefore, to quantify the perceptual effect of our approach, we conducted a user study on the Kodak dataset [7]. For each of the 24 uncompressed images in this dataset, the participants were asked to choose which of its two compressed versions looks better: the one with regular compression (with JPEG or JPEG 2000) or our deformation aware variant of the same compression method. For JPEG 2000, the two compared images were compressed with the same ratio (we tested ratios of 75, 125 and 175) and for JPEG both images were compressed with the same quality factor (we tested qualities of 20, 15 and 10). In the case of JPEG, our method resulted in a minor decrease in the compression ratio (~ 5% on average). The JPEG 2000 and JPEG surveys were completed by 57 and 59 Amazon Mechanical Turk workers, respectively. As can be seen in Fig. 14, the vast majority of the subjects chose our compressed images well above 50% of times. According to the Wilson test with confidence level of 95%, the mean percentage of preference for our method is $58.3\% \in [51.7\%, 79.9\%]$, $65.9\% \in [60.4\%, 87.6\%]$, $66\% \in [60.5\%, 87.7\%]$ for JPEG2000 with compression of ratios of 75, 125, 175, respectively, and $59.7\% \in [53.3\%, 80.8\%]$, $62.4\% \in [56.3\%, 83.5\%]$, $56.9\% \in [50.2\%, 78.0\%]$ for JPEG with quality indices of 20, 15, 25, respectively. This supports the conclusion that the average percentage of preference to our method is well above 50%, with very high confidence. This indicates that our deformation aware framework indeed leads to a meaningful improvement in visual quality over the original compression methods.

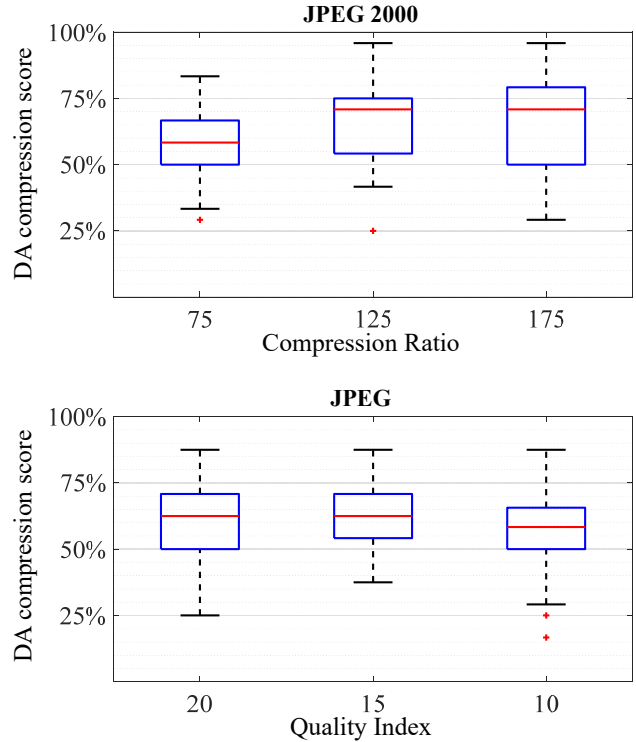


Figure 14. **User Study.** For each human subject, we recorded the percentage of times he/she preferred the image that was compressed with our deformation aware (DA) versions of JPEG 2000 and JPEG, over that which was compressed with the original methods (see text for details). For each compression ratio in JPEG 2000 and quality factor in JPEG, we plot the median percentage of preference (red line), the 25% and 75% percentiles (blue box), the extreme values (black lines), and outliers according to the interquartile ranges (IQR) (red marks). As can be seen, well above 75% of the subjects preferred our deformation aware version for more than 50% of the images.

6. Conclusions

We proposed a generic approach for improving the visual quality of lossy image compression schemes. Our method relies on a new error criterion, which is insensitive to smooth deformations. The advantages of our criterion are twofold. First, as opposed to other criteria, it can be easily incorporated into any existing compression scheme. Second, to excel under our criterion, the encoder need not invest bits in describing the exact geometries of fine structures. The effect is that more bits are invested in the important parts, leading to greatly better detail preservation. User studies confirmed that our approach significantly improves the visual quality of existing compression techniques.

Acknowledgements This research was supported in part by an Alon Fellowship, by the Israel Science Foundation (grant no. 852/17), and by the Ollendorf Foundation.

References

- [1] WebP: A New Image Format For The Web. <https://developers.google.com/speed/webp/>, 2015. 2, 6, 7
- [2] J. Ballé, V. Laparra, and E. P. Simoncelli. End-to-end optimized image compression. *International Conference on Learning Representations*, 2017. 8
- [3] M. F. Barnsley and A. D. Sloan. Methods and apparatus for image compression by iterated function system, 1990. US Patent 4,941,193. 1
- [4] F. Bellard. BPG image format. <https://bellard.org/bpg/>, 2014. 2, 6, 7
- [5] T. Dekel, T. Michaeli, M. Irani, and W. T. Freeman. Revealing and modifying non-local variations in a single image. *ACM Transactions on Graphics (TOG)*, 34(6), 2015. 3
- [6] P. Dollár and C. L. Zitnick. Structured forests for fast edge detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1841–1848, 2013. 6
- [7] R. Franzen. Kodak lossless true color image suite. <http://r0k.us/graphics/kodak>, year = 1999,. 6, 8
- [8] I. Galic, J. Weickert, M. Welk, A. Bruhn, A. Belyaev, and H.-P. Seidel. Towards PDE-based image compression. In *Variational, Geometric, and Level Set Methods in Computer Vision*, pages 37–48. 2005. 2
- [9] B. K. Horn and B. G. Schunck. Determining optical flow. *Artificial intelligence*, 17(1-3):185–203, 1981. 3, 4, 5
- [10] Y. Jiang and M. S. Pattichis. JPEG image compression using quantization table optimization based on perceptual image quality assessment. In *Asilomar Conference on Signals, Systems and Computers*, pages 225–229, 2011. 2, 3
- [11] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. *European Conference on Computer Vision*, 2016. 2
- [12] N. Johnston, D. Vincent, D. Minnen, M. Covell, S. Singh, T. Chinen, S. J. Hwang, J. Shor, and G. Toderici. Improved lossy image compression with priming and spatially adaptive bit rates for recurrent networks. *arXiv preprint arXiv:1703.10114*, 2017. 8
- [13] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. *arXiv:1609.04802*, 2016. 2
- [14] W. Lin and C.-C. J. Kuo. Perceptual visual quality metrics: A survey. *Journal of Visual Communication and Image Representation*, 22(4):297–312, 2011. 2
- [15] C. Liu. *Beyond pixels: exploring new representations and applications for motion analysis*. PhD thesis, 2009. 5
- [16] F. Luan, S. Paris, E. Shechtman, and K. Bala. Deep photo style transfer. *arXiv preprint arXiv:1703.07511*, 2017. 2
- [17] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A Database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *International Conference on Computer Vision*, volume 2, pages 416–423, July 2001. 6
- [18] M. Misiti, Y. Misiti, G. Oppenheim, and J.-M. Poggi. *Wavelets and their Applications*. John Wiley & Sons, 2013. 2, 6
- [19] B. Moghaddam, C. Nastar, and A. Pentland. A Bayesian similarity measure for direct image matching. In *International Conference on Pattern Recognition*, volume 2, pages 350–358, 1996. 3
- [20] T.-S. Ou, Y.-H. Huang, and H. H. Chen. SSIM-based perceptual rate control for video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 21(5):682–691, 2011. 3
- [21] T. Richter and K. J. Kim. A ms-ssim optimal jpeg 2000 encoder. In *Data Compression Conference*, pages 401–410, 2009. 2, 3
- [22] O. Rippel and L. Bourdev. Real-time adaptive image compression. In *International Conference on Machine Learning*, pages 2922–2930, 2017. 2, 8
- [23] T. Rott Shaham and T. Michaeli. Visualizing image priors. In *European Conference on Computer Vision*, pages 136–153. Springer, 2016. 2, 3
- [24] M. P. Sampat, Z. Wang, S. Gupta, A. C. Bovik, and M. K. Markey. Complex wavelet structural similarity: A new image similarity index. *IEEE transactions on image processing*, 18(11):2385–2401, 2009. 2
- [25] C. Schmaltz, P. Peter, M. Mainberger, F. Ebel, J. Weickert, and A. Bruhn. Understanding, optimising, and extending data compression with anisotropic diffusion. *International Journal of Computer Vision*, 108(3):222–240, 2014. 2
- [26] H. R. Sheikh and A. C. Bovik. Image information and visual quality. *IEEE Transactions on Image Processing*, 15(2):430–444, 2006. 2
- [27] H. R. Sheikh, A. C. Bovik, and G. De Veciana. An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Transactions on image processing*, 14(12):2117–2128, 2005. 2
- [28] A. Skodras, C. Christopoulos, and T. Ebrahimi. The JPEG 2000 still image compression standard. *IEEE Signal processing magazine*, 18(5):36–58, 2001. 1, 2, 6, 7
- [29] G. Strang and T. Nguyen. *Wavelets and filter banks*. SIAM, 1996. 2, 3, 6
- [30] L. Theis, W. Shi, A. Cunningham, and F. Huszár. Lossy image compression with compressive autoencoders. *International Conference on Learning Representations*, 2017. 8
- [31] G. Toderici, S. M. O’Malley, S. J. Hwang, D. Vincent, D. Minnen, S. Baluja, M. Covell, and R. Sukthankar. Variable rate image compression with recurrent neural networks. *arXiv preprint arXiv:1511.06085*, 2015. 8
- [32] G. Toderici, D. Vincent, N. Johnston, S. J. Hwang, D. Minnen, J. Shor, and M. Covell. Full resolution image compression with recurrent neural networks. *arXiv preprint arXiv:1608.05148*, 2016. 6, 7, 8
- [33] N. Wadhwa, T. Dekel, D. Wei, F. Durand, and W. T. Freeman. Deviation magnification: revealing departures from ideal geometries. *ACM Transactions on Graphics (TOG)*, 34(6):226, 2015. 3
- [34] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, and Y. Ma. Toward a practical face recognition system: Robust alignment and illumination by sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(2):372–386, 2012. 3

- [35] G. K. Wallace. The JPEG still picture compression standard. *IEEE transactions on consumer electronics*, 38(1):xviii–xxxiv, 1992. [1](#), [6](#), [7](#)
- [36] S. Wang, S. Ma, and W. Gao. SSIM based perceptual distortion rate optimization coding. In *Visual Communications and Image Processing*, pages 774407–774407. International Society for Optics and Photonics, 2010. [3](#)
- [37] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao. SSIM-motivated rate-distortion optimization for video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 22(4):516–529, 2012. [3](#)
- [38] Z. Wang and A. C. Bovik. Mean squared error: love it or leave it? A new look at signal fidelity measures. *IEEE signal processing magazine*, 26(1):98–117, 2009. [2](#)
- [39] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. [2](#)
- [40] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multiscale structural similarity for image quality assessment. In *Asilomar Conference on Signals, Systems and Computers*, volume 2, pages 1398–1402, 2004. [2](#)
- [41] L. Zhang, L. Zhang, X. Mou, and D. Zhang. FSIM: a feature similarity index for image quality assessment. *IEEE transactions on Image Processing*, 20(8):2378–2386, 2011. [2](#)
- [42] L. Zhang, L. Zhang, X. Mou, and D. Zhang. A comprehensive evaluation of full reference image quality assessment algorithms. In *IEEE International Conference on Image Processing*, pages 1477–1480, 2012. [2](#)