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.*

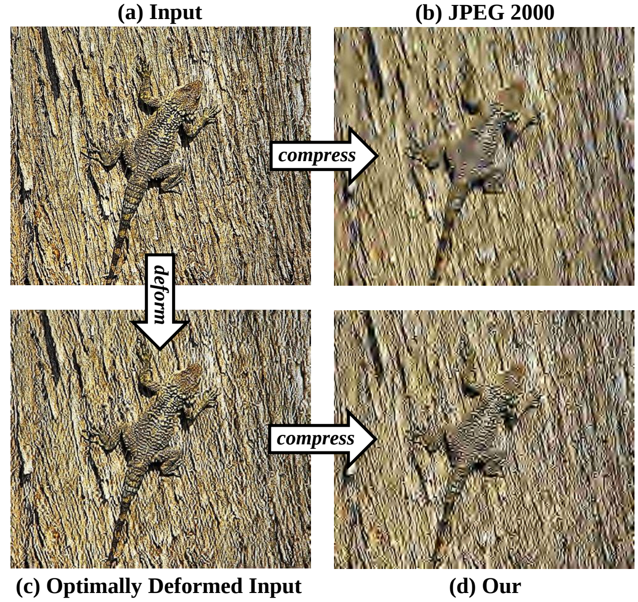

Figure 1. **Deformation aware image compression.** Our algorithm seeks to minimize a deformation-insensitive error measure. This boils down to determining how to best deform the input image (a) so as to make it more compressible (c). By doing so, we trade a little geometric integrity with a significant gain in terms of preservation of visual information (d) compared to regular compression (b).

1. Introduction

High-resolution cameras have become extremely popular over the last two decades (*e.g.* in mobile devices). To accommodate the numerous amounts of pictures captured by such devices, high quality lossy compression algorithms are a necessity. In this work, 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 algorithm uses the CODEC as a "black box" and is thus very 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.

Most compression methods seek to minimize some per-pixel distance (typically ℓ_2) between the input image and the decoded image. We claim that the main limitation of such distance measures 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, which is clearly wasteful, as the human visual system is not sensitive to small geometric deformations as long as the semantics of the scene is preserved.

Motivated by this insight, 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

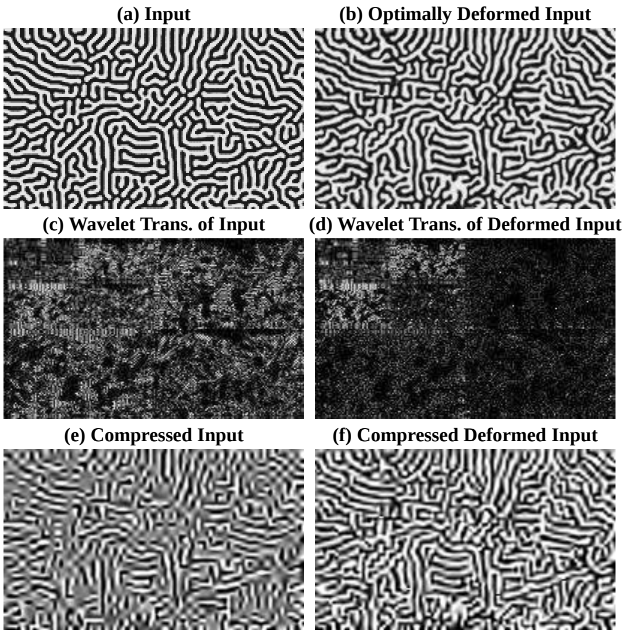

Figure 2. **Deformation aware compression via Subband Thresholding.** The input image (a) contains many strong curved edges, so that its wavelet transform (c) is not very sparse. This causes Subband Thresholding compression [8] at a ratio of 25:1, to produce a very blurry result (e). However, by introducing 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 any image priors.

under our criterion boils down to determining how to best deform the input image such that it becomes more "compressible". 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. This is illustrated in Fig. 1.

The surprising success of our approach can be attributed to an interesting phenomenon recently observed in [6]. That is, by introducing small 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. 2 for the Subband Thresholding compression method [8]. However, as can be seen in Fig. 2(b),(d), it takes only a small deformation to make the wavelet transform of the image significantly sparser. Thus, by slightly sacrificing geometric integrity, we substantially improve the ability of the compression algorithm to preserve details, as seen in Fig. 2(f).

2. 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

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

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 $d_{\text{SSD}}(x,y) = \|x-y\|^2$. The SSD is a per-pixel criterion, and is therefore extremely sensitive to slight misalignment or deformation of objects. Therefore, when the bit budget ε is low (*i.e.* high compression ratio), the encoder severely blurs the fine structures in the image.

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

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

where the term $\psi(\mathcal{T})$ penalizes for non-smooth deformations. In other words, DASSD is 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

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

We define the penalty $\psi(\mathcal{T})$ to be a weighted Horn and Schunk regularizer [4].

To see why deformation invariance improves compressions, note that lossy 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. 3 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

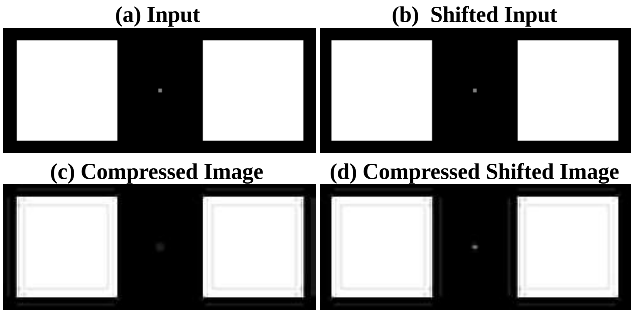

Figure 3. **The effect of global translation.** Compressing the image (a) using JPEG 2000 at a ratio of 75:1, causes the small square in the middle to disappear (c). However, by shifting the image only two pixels to the left (b), compression at the same ratio keeps the small square intact (d).

criterion prefers the result in which the small square is preserved: The DASSD between (a) and (d) is 3% lower 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).

Algorithm

Substituting d_{DASSD} of (2) into (1), we obtain the optimization problem

$$\min_{x,\mathcal{T}} \|\mathcal{T}\{y\} - x\|^2 + \lambda\,\psi(\mathcal{T}) \quad \text{s.t.} \quad R(x) \le \varepsilon. \qquad (3)$$

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

To solve problem (3) we alternate between minimizing the objective w.r.t. x while holding $\mathcal{T}$ fixed and then w.r.t. $\mathcal{T}$ while holding x fixed. This amounts to compressing the current deformed input image $\mathcal{T}\{y\}$ to obtain x, and computing the optical flow between x and y to update $\mathcal{T}$.

3. Experiments

We tested our approach with JPEG [10], JPEG 2000 [7], WebP [1], BPG [2], and the deep-net based CODEC of [9]. Figures 1 and 4 show several results produced by our algorithm. As can be seen, our algorithm manages to preserve a lot of the content that is otherwise completely lost in regular compression.

To quantify the perceptual effect of our approach when used with JPEG and JPEG 2000, we conducted a user study on the Kodak dataset [3]. 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 the one with our deformation aware variant of the same compression method. The results are summarized in Fig. 5. As can be seen, the vast majority of the subjects chose our compressed images well above 50% of times. This indicates that our deformation aware framework leads to a significant improvement in visual quality over the original compression methods.

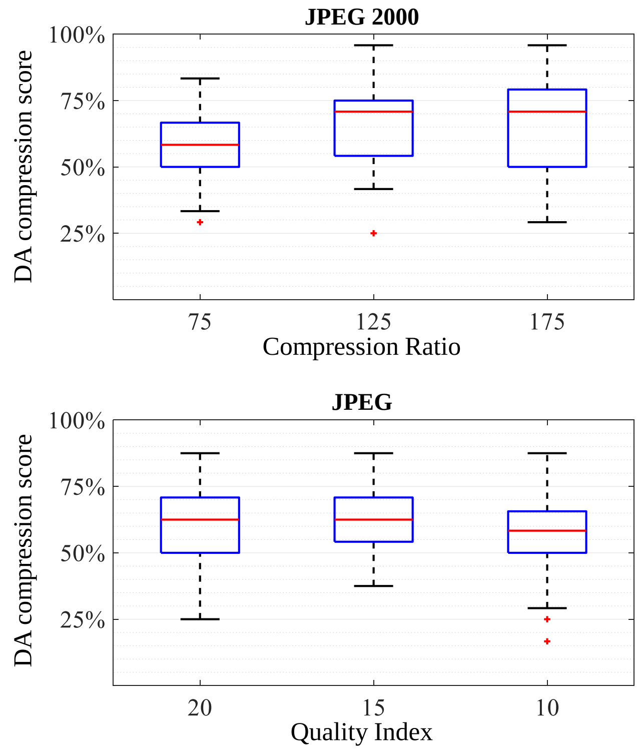

Figure 5. For each human subject, we recorded the percentage of times he/she preferred the image that was compressed with our deformation aware (DA) versions of of JPEG 2000 and JPEG, over that which was compressed with the original methods. For each compression ratio in JPEG 2000 and quality factor 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.

References

[1] WebP: A New Image Format For The Web. https://developers.google.com/speed/webp/, 2015. 3

[2] F. Bellard. BPG image format. https://bellard.org/bpg/, 2014. 3

[3] R. Franzen. Kodak lossless true color image suite. http://r0k.us/graphics/kodak, 1999. 3

[4] B. K. Horn and B. G. Schunck. Determining optical flow. *Artificial intelligence*, 17(1-3):185–203, 1981. 2

[5] C. Liu. *Beyond pixels: exploring new representations and applications for motion analysis*. PhD thesis, 2009.

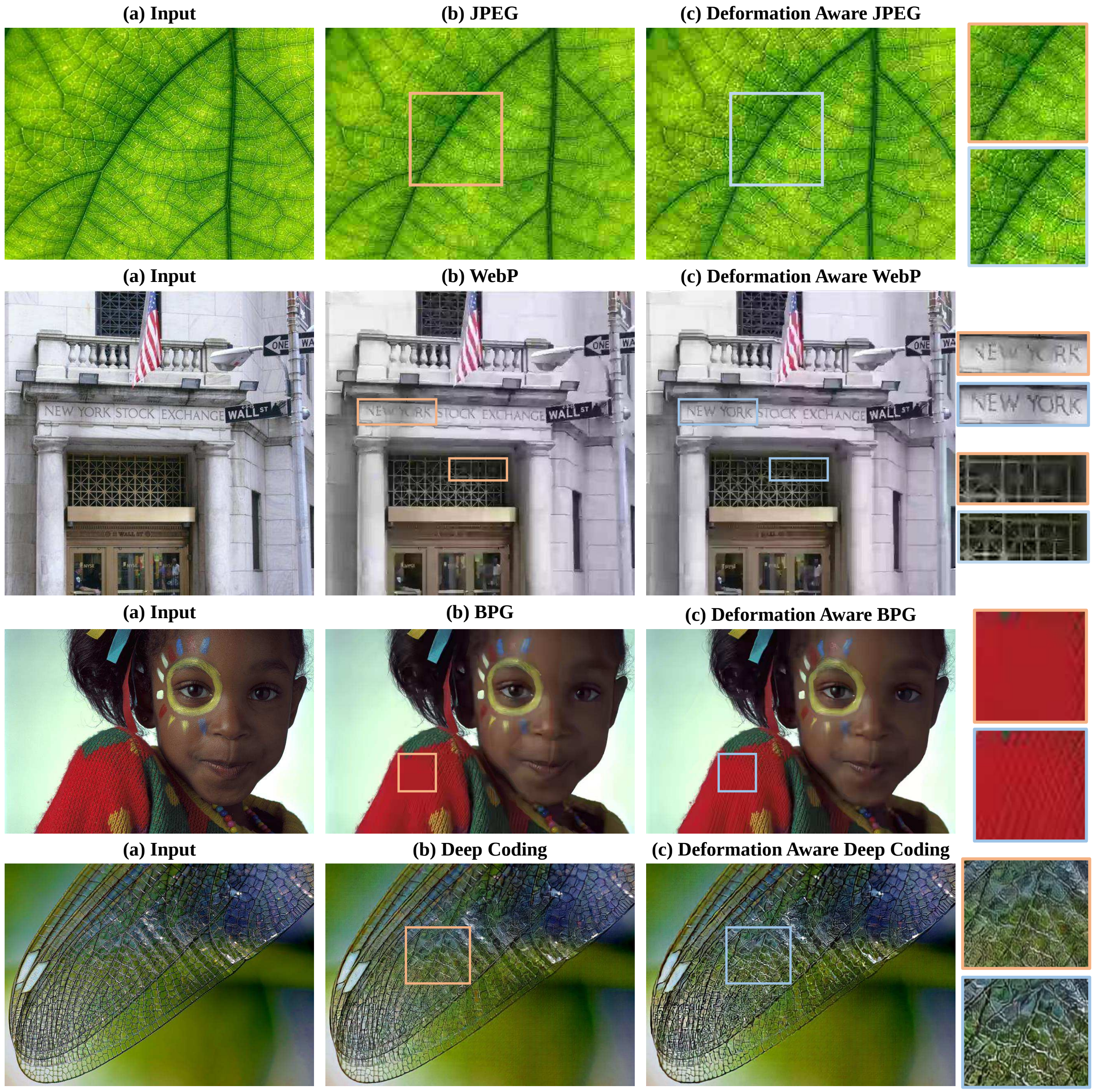

Figure 4. **Visual results.** We compare the results of the tested methods with our Deformation Aware variant. Top row to bottom: JPEG at a ratio of 50:1, WebP at a ratio of 110:1, BPG at a ratio of 220:1 and Deep coding at a ratio of 48:1.

[6] T. Rott Shaham and T. Michaeli. Visualizing image priors. In *European Conference on Computer Vision*, pages 136–153. Springer, 2016. 2

[7] A. Skodras, C. Christopoulos, and T. Ebrahimi. The JPEG 2000 still image compression standard. *IEEE Signal processing magazine*, 18(5):36–58, 2001. 3

[8] G. Strang and T. Nguyen. *Wavelets and filter banks*. SIAM, 1996. 2

[9] 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. 3

[10] G. K. Wallace. The JPEG still picture compression standard. *IEEE transactions on consumer electronics*, 38(1):xviii–xxxiv, 1992. 3