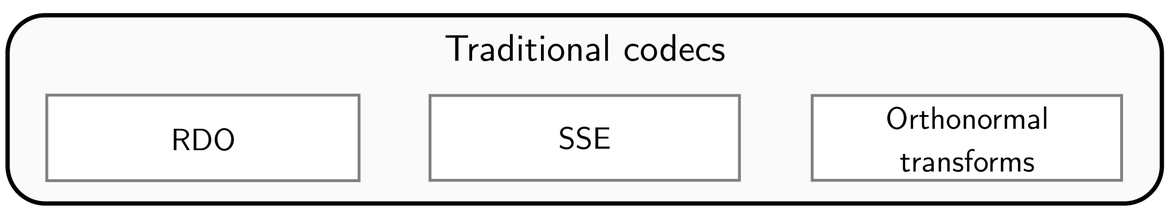
FEATURE-PRESERVING RDO IN IMAGE CODING FOR MACHINES

USC Viterbi School of Engineering
Ming Hsieh Department of Electrical and Computer Engineering

Samuel Fernández **Eduardo Pavez** **Antonio Ortega**

Motivation



- RDO: Optimal codec parameters based on the input.
- SSE is localized $\implies$ aggregation of block-wise SSE.
- Orthonormal transforms: by Parseval, RDO in transform domain.

Problem: SSE is adequate for the human visual system.

But many images and videos are used to extract semantic information.

Coding for machines: *compress while preserving task performance* [BXO03].

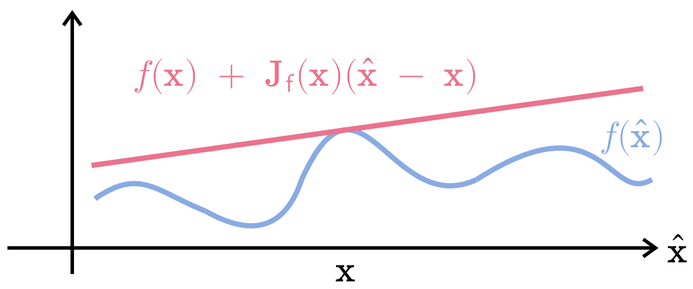
Metric linearization

We can apply Taylor's expansion to the feature extractor around $\mathbf{x}$:

$$f(\hat{\mathbf{x}}(\theta)) = f(\mathbf{x}) + \mathbf{J}_\mathrm{f}(\mathbf{x})(\hat{\mathbf{x}} - \mathbf{x}) + o(||\hat{\mathbf{x}} - \mathbf{x}||_2^2). \tag{1}$$

Thus, the feature-distortion becomes

$$||f(\mathbf{x}) - f(\hat{\mathbf{x}})||_2^2 \approx ||\mathbf{J}_\mathrm{f}(\mathbf{x})(\hat{\mathbf{x}} - \mathbf{x})||_2^2. \tag{2}$$



Metric localization

Still not suitable for block-wise evaluation. Localize the metric:

$$||\mathbf{J}_\mathrm{f}(\mathbf{x})(\hat{\mathbf{x}} - \mathbf{x})||_2^2 \approx \sum_{i=1}^{n_b} ||\mathbf{J}_\mathrm{f}^{(i)}(\mathbf{x})(\hat{\mathbf{x}}_i - \mathbf{x}_i)||_2^2. \tag{3}$$

Thus, we obtain an input-dependent squared error (IDSE) loss:

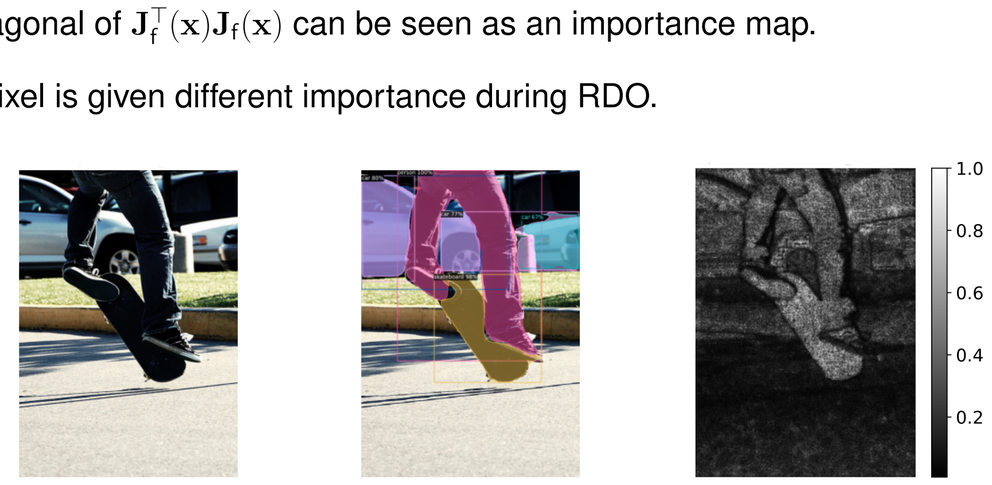
$$\theta_i^\star = \arg\min_{\theta\in\Theta} ||\mathbf{J}_\mathrm{f}^{(i)}(\mathbf{x})(\hat{\mathbf{x}}_i - \mathbf{x}_i)||_2^2 + \lambda\, r_i(\hat{\mathbf{x}}_i(\theta)). \tag{4}$$

The Jacobian measures the influence of each pixel on the features.

Importance maps

The diagonal of $\mathbf{J}_\mathrm{f}^\top(\mathbf{x})\mathbf{J}_\mathrm{f}(\mathbf{x})$ can be seen as an importance map.

Each pixel is given different importance during RDO.



Jacobian approximation and SSE

Using linear dimensionality reduction:

$$\mathbf{J}_{\mathrm{S}\mathrm{f}}(\mathbf{x}) = \mathbf{S}\mathbf{J}_\mathrm{f}(\mathbf{x}). \tag{5}$$

We reduce the dimensionality of the features before computing the Jacobian.

Johnson–Lindenstrauss [Ach03]: if entries of $\mathbf{S}$ are i.i.d.,

$$(1-\epsilon)||\mathbf{S}\mathbf{J}_\mathrm{f}^{(i)}(\mathbf{x})(\hat{\mathbf{x}}_i - \mathbf{x}_i)||_2^2 \le ||\mathbf{J}_\mathrm{f}^{(i)}(\mathbf{x})(\hat{\mathbf{x}}_i - \mathbf{x}_i)||_2^2 \le (1+\epsilon)||\mathbf{S}\mathbf{J}_\mathrm{f}^{(i)}(\mathbf{x})(\hat{\mathbf{x}}_i - \mathbf{x}_i)||_2^2.$$

The result can also be combined with the SSE:

$$(\mathbf{x}_i - \hat{\mathbf{x}}_i)^\top \left(\mathbf{J}_\mathrm{f}^{(i)}(\mathbf{x})^\top \mathbf{S}^\top \mathbf{S}\mathbf{J}_\mathrm{f}^{(i)}(\mathbf{x}) + \tau\,\mathbf{I}\right)(\mathbf{x}_i - \hat{\mathbf{x}}_i). \tag{6}$$

Tikhonov regularization: τ controls minimum SSE for a given pixel.

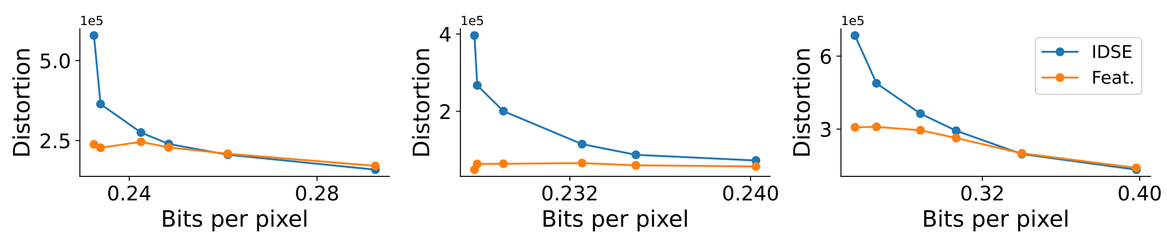
Feature-based RDO

Assume we have access to a *relevant* feature extractor $f(\cdot)$.

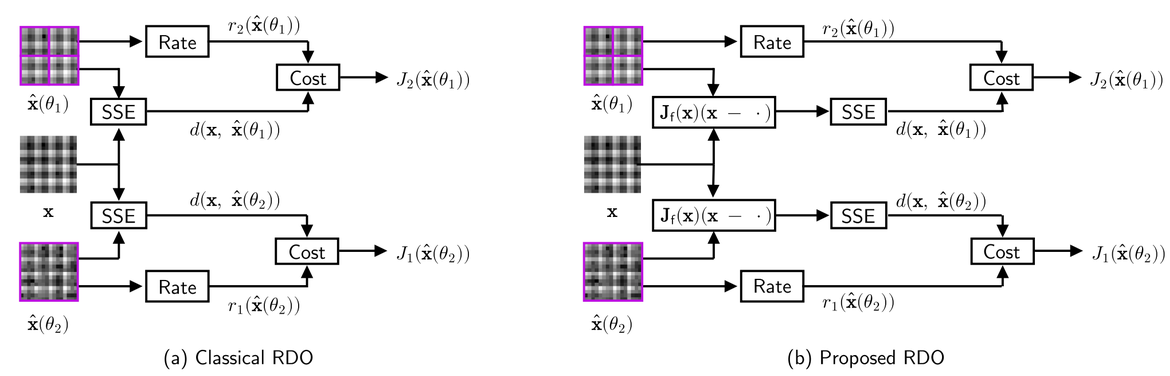
Compress to preserve features as much as possible [Fis+20]:

$$\theta^\star = \arg\min_{\theta\in\Theta} ||f(\mathbf{x}) - f(\hat{\mathbf{x}}(\theta))||_2^2 + \lambda \sum_{i=1}^{n_b} r_i(\mathbf{x}_i(\theta)). \tag{7}$$

Challenges: decisions are made block-wise.



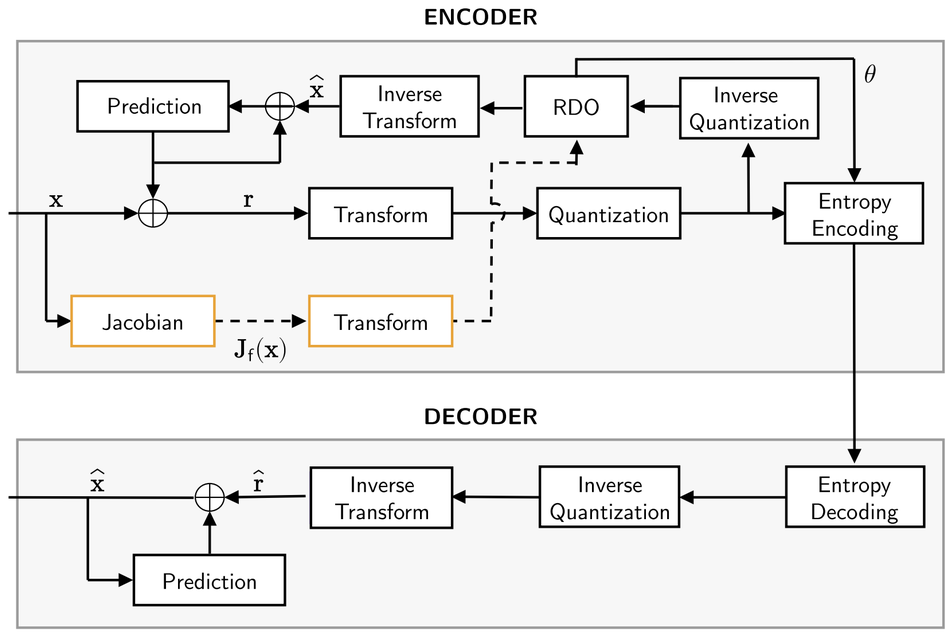
IDSE-RDO



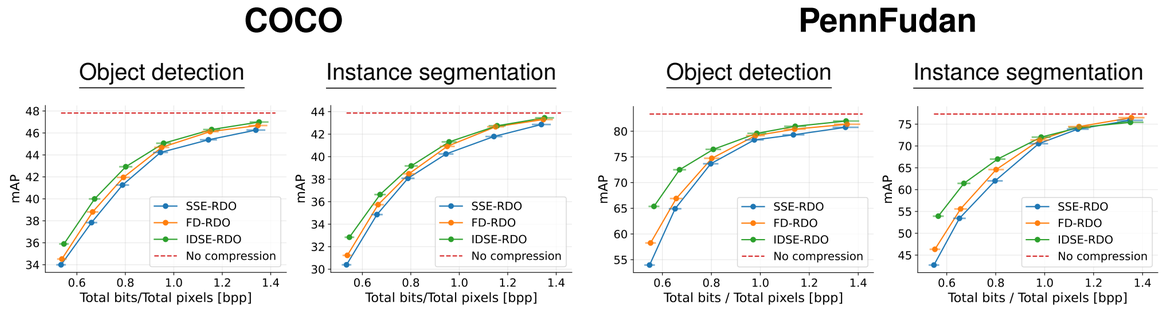
(a) Classical RDO (b) Proposed RDO

Caveat: the Jacobian is very high dimensional.

Standard compliant encoder



Empirical results



Dims.	Time [s]	PSNR BD-R. [%]	mAP(D) BD-R. [%]	mAP(S) BD-R. [%]
$\ell = 2$	0.067	1.12	−6.31	−6.01
$\ell = 4$	0.112	0.82	−7.18	−7.06
$\ell = 8$	0.212	0.79	−8.28	−7.77

Original dimension: 1,066,240. For BD-rate, more negative is better.

References

[Ach03] Dimitris Achlioptas. "Database-friendly random projections: Johnson-Lindenstrauss with binary coins". In: *Journal of Comput. and Sys. Sciences* 66.4 (2003), pp. 671–687.

[BXO03] B. Beferull-Lozano, Hua Xie, and A. Ortega. "Rotation-invariant features based on steerable transforms with an application to distributed image classification". In: *Proc. IEEE Int. Conf. Image Process.* Vol. 3. 2003, pp. 517–521.

[Fis+20] Kristian Fischer et al. "Video Coding for Machines with Feature-Based Rate-Distortion Optimization". en. In: *Proc. IEEE Int. Work. Mult. Signal Process.* IEEE, Sept. 2020, pp. 1–6. ISBN: 978-1-72819-320-5.