Lucas-Kanade Reloaded:
End-to-End Super-Resolution from Raw Image Bursts

Julien Mairal

Inria Grenoble
Collaborators
with a picture of me because my webcam is broken

Bruno Lecouat
Jean Ponce
me (five years ago)

A 20-megapixel innocent scene
...taken at high ISO with low exposure time

Left: high-quality jpg output of the camera ISP.
...taken at high ISO with low exposure time

Left: high-quality jpg output of the camera ISP.
Right: $\times 4$ super-resolution, after processing a burst of 30 raw images (handheld camera).
...taken at high ISO with low exposure time

Left: high-quality jpg output of the camera ISP.
Right: \( \times 4 \) super-resolution, after processing a burst of 30 raw images (handheld camera).
...taken at high ISO with low exposure time

Left: high-quality jpg output of the camera ISP.
Right: ×4 super-resolution, after processing a burst of 30 raw images (handheld camera).
The Camera raw processing pipeline (simplified view)

How does your camera process sensor data?


Conversion to sRGB. Gamma correction.

Denoising

Demosaicking
The Camera raw processing pipeline (simplified view)

How does your camera process sensor data?


Conversion to sRGB. Gamma correction.

Denoising

Demosaicking

Idea: working with raw data is important, before the camera ISP produces irremediable damage!
With raw data, we may leverage aliasing!

Figure: Example of aliasing: undersampled sinusoid causes confusion with a sinusoid with lower frequency. Picture from Wikipedia.

- Aliasing is usually mitigated with some optical / digital filters.
- If we analyze the aliasing patterns from multiple frames we can recover high frequencies.
Super-resolution from raw image bursts (with natural hand motion)

This is hard because it requires, simultaneously,

- accurately aligning images with subpixel accuracy.
- dealing with noisy data (blind denoising).
- reconstructing color images from the Bayer pattern (demosaicking).
Multiframe super resolution: prior work

and, among many others:

- **interpolation-based methods**: [Hardie, 2007], [Takeda et al., 2007];
- **iterative approaches**: [Irani and Peleg, 1991], [Elad and Feuer, 1997], [Farsiu et al., 2004];
- **(deep) learning-based approaches**: [Bhat et al., 2021], [Molini et al., 2019], [Deudon et al., 2019];
- and also the literature on video super-resolution (typically not dealing with raw data).

**Interesting for us: synthetic raw datasets from Bhat et al. [2021].**
The “old” world of classical inverse problems.

Image formation model

\[ y_k = DBW_{p_k} x + \varepsilon_k. \]

Inverse problem given \( y_1, \ldots, y_K \)

\[
\min_{x,p_k} \frac{1}{K} \sum_{k=1}^{K} \| y_k - \underbrace{DBW_{p_k} x}_{U_{p_k}} \|_2^2 + \lambda \phi_\theta(x).
\]

A natural strategy

- define an appropriate prior \( \phi_\theta(x) \) for natural images and optimize!
The “old” world of classical inverse problems.

Simple relaxation with “half quadratic splitting” + block coordinate descent

\[
\min_{x,z,p_k} \frac{1}{K} \sum_{k=1}^{K} \|y_k - U_{p_k}z\|^2 + \frac{\mu_t}{2} \|z - x\|^2 + \lambda \phi_\theta(x).
\]

- minimizing with respect to \(p_k\) (parameters of an affine transformation) is performed by Gauss-Newton steps. This is the algorithm of Lucas and Kanade [1981].
- minimizing with respect to \(x\) requires computing the proximal operator of \(\phi_\theta\).
- minimizing w.r.t. \(z\) can be done by gradient descent steps.
- \(\mu_t\) increases over the iterations.
The “old” world of classical inverse problems.

Simple relaxation with “half quadratic splitting” + block coordinate descent

\[
\min_{x,z,p_k} \frac{1}{K} \sum_{k=1}^{K} \|y_k - U_{p_k} z\|^2 + \frac{\mu_t}{2} \|z - x\|^2 + \lambda \phi_\theta(x).
\]

- minimizing with respect to \(p_k\) (parameters of an affine transformation) is performed by Gauss-Newton steps. This is the algorithm of Lucas and Kanade [1981].
- minimizing with respect to \(x\) requires computing the proximal operator of \(\phi_\theta\).
- minimizing w.r.t. \(z\) can be done by gradient descent steps.
- \(\mu_t\) increases over the iterations.

**Advantage:** robustness and interpretability (solves what it is supposed to solve).

**Drawback:** designing a good image prior by hand is hard.
The “new” world of deep learning models (Pic. https://xkcd.com/)

- a form of prior knowledge is encoded in the model architecture (e.g., a convolutional neural network for images).
- ability to train model parameters $\theta$ end to end.
- state-of-the-art for many tasks (once the right model/setup is found).
- requires training data.

**Advantage:** task-adaptive.
**Drawback:** tuned to specific data distribution.
Bridging the two worlds with trainable algorithms.

Idea 1: plug-and-play priors [Venkatakrishnan et al., 2013]
Replace proximal operator
\[
\arg\min_x \frac{1}{2} \|z - x\|^2 + \lambda \phi_\theta(x),
\]
by a convolutional neural network \(f_\theta(z)\).
Bridging the two worlds with trainable algorithms.

**Idea 1: plug-and-play priors [Venkatakrishnan et al., 2013]**

Replace proximal operator

\[
\arg \min_x \frac{1}{2} \| z - x \|^2 + \lambda \phi_\theta(x),
\]

by a convolutional neural network \( f_\theta(z) \).

**Idea 2: unrolled optimization [Gregor and LeCun, 2010]**

- Consider the previous optimization procedure with \( T \) steps, producing an estimate \( \hat{x}_T(Y) \), given a burst \( Y = y_1, \ldots, y_K \).
- Given a dataset of training pairs \( (x_i, Y_i)_{i=1, \ldots, n} \), minimize

\[
\min_{\theta} \frac{1}{n} \sum_{i=1}^{n} \| \hat{x}_T(Y_i) - x_i \|_1.
\]
we keep the interpretability of the classical inverse problem formulation.
we benefit from a data-driven image prior.
Extreme $\times 16$ super-resolution.

Figure: Experiment with a synthetic RGB burst of 20 images with random affine motions.
Experiments on real raw data - Pixel 4a.

Figure: Full scene - camera ISP - Our $\times 4$ results.
Experiments on real raw data - Pixel 4a.

Figure: Full scene - camera ISP - Our $\times 4$ results.
Current issues with moving objects

Figure: Misalignments artefacts due to moving objects in the scene. Our current implementation does not handle fast moving objects and then generates visual artefacts.
Conclusion

Take-home messages

- 40-years old computer vision algorithms are useful.
- aliasing is good.
- “classical” approaches are robust and interpretable and greatly benefit from deep learning principles (differentiable programming).

Future work

- microscopy and astronomical imaging where we want to recover “true” signals.
- high-quality and high-dynamic range panoramas.
- going beyond static scenes.


References II


