Computation + Photography
How the mobile phone became a camera

Peyman Milanfar
Mobile imaging has changed the world.
Vatican Square

2005
Pope Benedict announcement

2013
Pope Francis announcement
More than 2 billion photos shared on social media per day

That’s 23,000 frames/sec

Over 100 million are “selfies”

That’s 1,200 frames/sec
Smartphones Cause Photography Boom

Number of digital photos taken worldwide*

<table>
<thead>
<tr>
<th>Year</th>
<th>Photos Taken (in billions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013</td>
<td>660b</td>
</tr>
<tr>
<td>2014</td>
<td>810b</td>
</tr>
<tr>
<td>2015</td>
<td>1,000b</td>
</tr>
<tr>
<td>2016</td>
<td>1,100b</td>
</tr>
<tr>
<td>2017</td>
<td>1,200b</td>
</tr>
</tbody>
</table>

Devices used in 2017

- 85.0% Smartphones
- 10.3% Digital cameras
- 4.7% Tablets

Source: Statista
What Smartphones Have Done to the Camera Industry

Worldwide shipments of photo cameras by CIPA members since 1951*

Source: Statista
Instant Gratification

Circa 1978
Year

First Digital Camera Prototype
First Commercial Digital Cameras
CMOS
Invention of CMOS/Camera on a Chip

It would take another 10 years before CMOS systems would enable mass production of affordable (mobile) cameras

+ Cheaper, power efficient
- Noisier, rolling shutter readout

MOBILE PHOTOGRAPHY
First Digital Camera Prototype

First Commercial Digital Cameras

1st Commercial Camera Phone

Digital SLRs, Compacts

CMOS

iPhone
Displays

1990

2010

2020

300dpi displays
First Digital Camera Prototype

First Commercial Digital Cameras

1st Commercial Camera Phone

4G Networks

- Designed Primarily for Data
- IP based protocol
- True Mobile broadband.

Digital SLRs, Compacts

CMOS

iPhone
Wireless Network Speed

Transmit a 2 Mpix image

- 30 seconds
- 0.5 seconds
- 0.01 seconds

Source: Silika
2010 -

COMPUTATIONAL PHOTOGRAPHY

“The best camera is the one that’s with you.”
Can one be as good as the other?
Can one be as good as the other?

4.7 mm²

~ 300x =

1360 mm²
Less light gets recorded
Compete with hardware!

Yet most of the improvements are due to software.
Modern Mobile Imaging: Burst Photography

Exposure control

**Align:** Reliable Optical Flow – Scene is never stationary

**Merge:** Artifact-free Fusion – Alignment failures, occlusion, ...

**Enhance:** Denoise, Sharpen, Contrast, Dynamic Range
Classic Camera Image Processing Pipeline

Sensor with color filter array (CCD/CMOS) → Gain Control A/D Converter Possible LUT → White Balance

“Enhance”

- Tone Reproduction
- Color Space Transform + Color Preferences
- Noise Reduction/Sharpening

“Merge”

- Demosaic
- JPEG Compression
Classic camera pipeline - demosaicing

Two-thirds of your picture is made-up!
Demosaiicing
Demosaicing .... Kills Details and Produces Artifacts
Instead ..... Replace demosaicing with multiple frames
How: “Pixel-shifting”

Shift sensor right 1 pixel

Shift sensor down 1 pixel

Shift sensor down and right 1 pixel
Life is not so simple.
Multi-dimensional, non-uniform, interpolation
Merge: Nonlinear Kernel Regression

Continuous interpolation

Measurements

Kernel functions
We can also merge onto higher-res grid

- This has its limits
  - depending on the pixel size / lens spot size tradeoff
  - for typical mobile sensors, up to ~2x is possible
Source of motion in mobile imaging?
Handheld burst capture
(Natural) Physiological Tremor

Measured in 100s of bursts
What if phone/camera is immobilized?

Simulated “tremor”
Motion : Phase Diversity

Aliasing + Phase diversity $\rightarrow$ Multi-frame Super-Res

Aliasing + Subpixel Motion

Super-res
The visual system appears to do super-resolution (via micro-saccades)

Ours eyes are constantly in motion. Even during visual fixation, small eye movements continually jitter the location of gaze\(^1\). It is known that visual percepts tend to fade when retinal image motion is eliminated in the laboratory\(^2\). However, it has long been debated whether, during natural viewing, fixational eye movements have functions in addition to preventing the visual scene from fading\(^3\). In this study, we analyzed the influence in humans of fixational eye movements on the discrimination of gratings masked by noise that has a power spectrum similar to that of natural images. Using a new method of retinal image stabilization\(^4\), we selectively eliminated the motion of the retinal image that normally occurs during the intersaccadic intervals of visual fixation. Here we show that fixational eye movements improve discrimination of high spatial frequency stimuli, but not of low spatial frequency stimuli. This improvement originates from the temporal modulations introduced by fixational eye movements in the visual input to the retina, which emphasize the high spatial frequency harmonics of the stimulus. In a natural visual world dominated by low spatial frequencies, fixational eye movements appear to constitute an effective sampling strategy by which the visual system enhances the processing of spatial detail.

stabilization during periods of visual fixation between saccades, as would have been necessary to study fixational eye movements in their natural context\(^5\). Instead, all trials with stabilized vision had to be run in uninterrupted blocks while the subject maintained fixation—a highly unnatural condition that unavoidably led to visual fatigue and fading.

In this study, we examined the influence of fixational eye movements on the discrimination of targets at different spatial frequencies (grating spacings). We compared discrimination performances measured in two conditions: with and without the retinal image motion produced by fixational eye movements. To overcome the limitations of previous experiments, we developed a new retinal stabilization technique based on real-time processing of eye-movement signals\(^6\). Like previous stabilization methods, this technique does not guarantee perfect elimination of retinal image motion; however, unlike previous methods, it combines a high quality of stabilization with experimental flexibility (see Supplementary Information). This flexibility enabled us to display and selectively stabilize the stimulus after a saccade, a method that isolates the normal fixational motion of the eye. It also allowed us to randomly alternate between trials with retinal stabilization and trials with normal retinal motion, a procedure that...
Crops

Hasinoff et al. [2016]  Ours
"The Pixel 3 is the first smartphone camera to truly challenge traditional cameras from an image quality standpoint, . . . rivaling cameras with Micro 4/3 sensors in [super-res] mode."
[SIGGRAPH 2019]

Handheld Multi-Frame Super-Resolution

BARTLOMIEJ WRONSKI, IGNACIO GARCIA-DORADO, MANFRED ERNST, DAMIEN KELLY, MICHAEL KRAININ, CHIA-KAI LIANG, MARC LEVOY, and PEYMAN MILANFAR, Google Inc.
Zoom Use Case

Align → Merge → Enhance → Crop → Upscale

Enhance! RAISR Sharp Images with Machine Learning
Monday, November 14, 2016

Posted by Peyman Milanfar, Research Scientist

[Romano, Milanfar, Isidoro, Transactions on Computational Imaging, 2017]
Filter Learning

Low res images \rightarrow \text{patches} \rightarrow \text{least-squares solver} \rightarrow \text{filter} \rightarrow \text{pixels}

\[
\min_h \| Ah - b \|^2_2
\]
We can do even better

- Bucket similar patches together and train within buckets

LR images

cheap upscaling

hashing to buckets

hash=0 patches

hash=1 patches

hash=2 patches

hash=n patches

filter learning

hash=0 pixels

hash=1 pixels

hash=2 pixels

hash=n pixels

Learning per bucket

HR images

hashing to buckets
Learned 2x Upscaling Filters
No zoom
(2x zoom crop) Standard Digital Zoom
(2x zoom crop) 2017 Single-frame Super-res
(2x zoom crop)

2018 Multi-frame Super-res
85% of optical zoom resolution at 2x

“Best digital zoom on the market”
OTHER CHALLENGES IN COMPUTATIONAL IMAGING
Curation

Introducing NIMA: Neural Image Assessment
Monday, December 18, 2017

Posted by Hossein Talebi, Software Engineer and Peyman Milanfar Research Scientist, Machine Perception

Quantification of image quality and aesthetics has been a long-standing problem in image processing and computer vision. While technical quality assessment deals with measuring pixel-level degradations such as noise, blur, compression artifacts, etc., aesthetic assessment captures semantic level characteristics associated with emotions and beauty in images. Recently, deep convolutional neural networks (CNNs) trained with human-labelled data have been used to address the subjective nature of image quality for specific classes of images, such as landscapes. However, these approaches can be limited in their scope, as they typically categorize images to two classes of low and high quality. Our proposed method predicts the distribution of ratings. This leads to a more accurate quality prediction with higher correlation to the ground truth ratings, and is applicable to general images.

NIMA: Neural Image Assessment

NIMA for **Aesthetic Quality**
NIMA For Technical Quality
Shot on Pixel 3
With Super-res
Night Sight mode