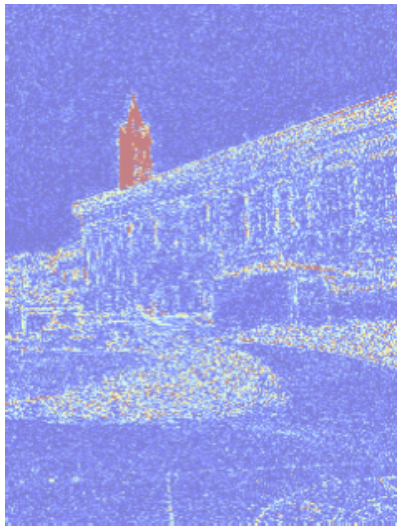
Impeding Forgers at Photo Inception

Matthias Kirchner [a]
Peter Winkler [b]
Hany Farid [b]

[a] ICSI Berkeley | [b] Dartmouth College





Media Watermarking, Security, and Forensics

Burlingame, CA | February 5, 2013

Information matters

low-res grayscale



full-res color



stereo vision



ease of image forgery

available scene information

- varied set of forensic techniques for authenticating digital images
- the more information the camera records the more difficult it is to create a compelling fake image

Information matters

low-res grayscale



full-res color



stereo vision



ease of image forgery

available scene information

lossy post-processing impedes forensics

- varied set of forensic techniques for authenticating digital images
- the more information the camera records the more difficult it is to create a compelling fake image

additional information impedes forgers

Information matters

low-res grayscale



full-res color



stereo vision

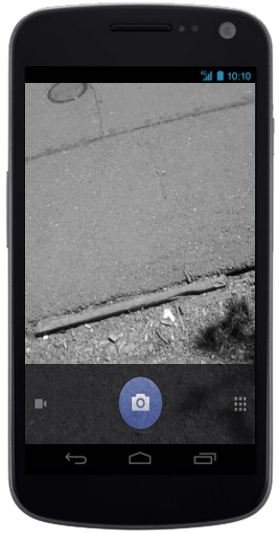


ease of image forgery

available scene information

standard consumer devices ?

- most devices provide a digital preview
  - ▷ if recorded with the image, manipulations must propagate
- motion sensors
  - ▷ 3-D structure of a forged preview must be consistent with recorded camera motion



Frame alignment procedure

- record and store N preview frames prior to image capture (≈10 fps, 320 × 240, JPG 85)

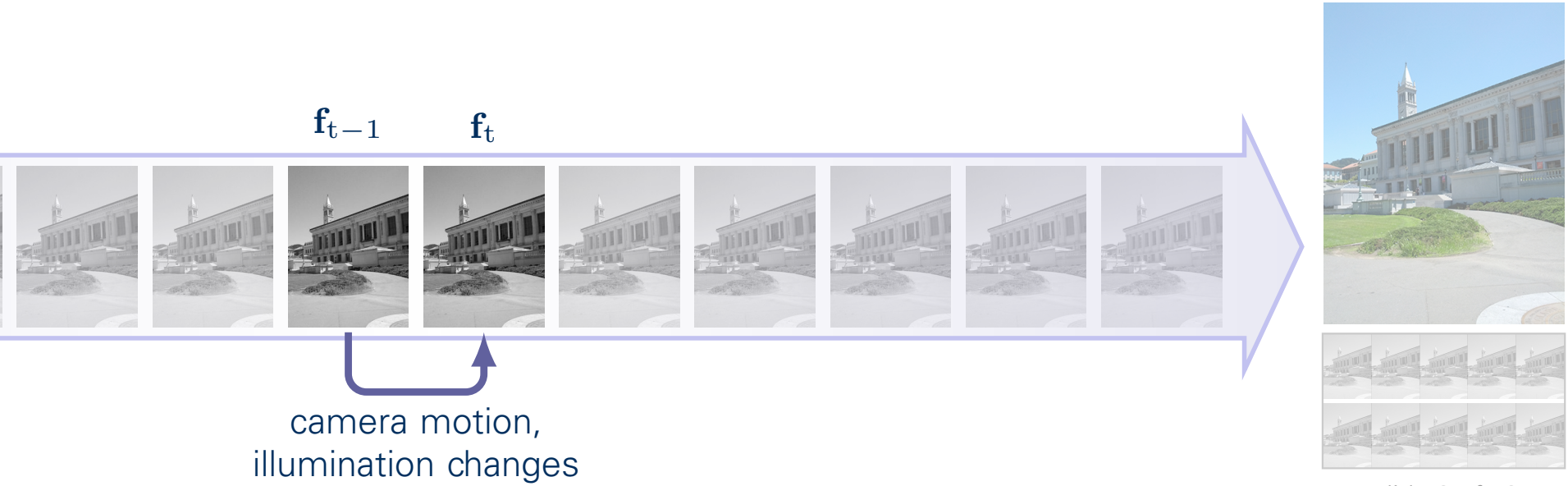






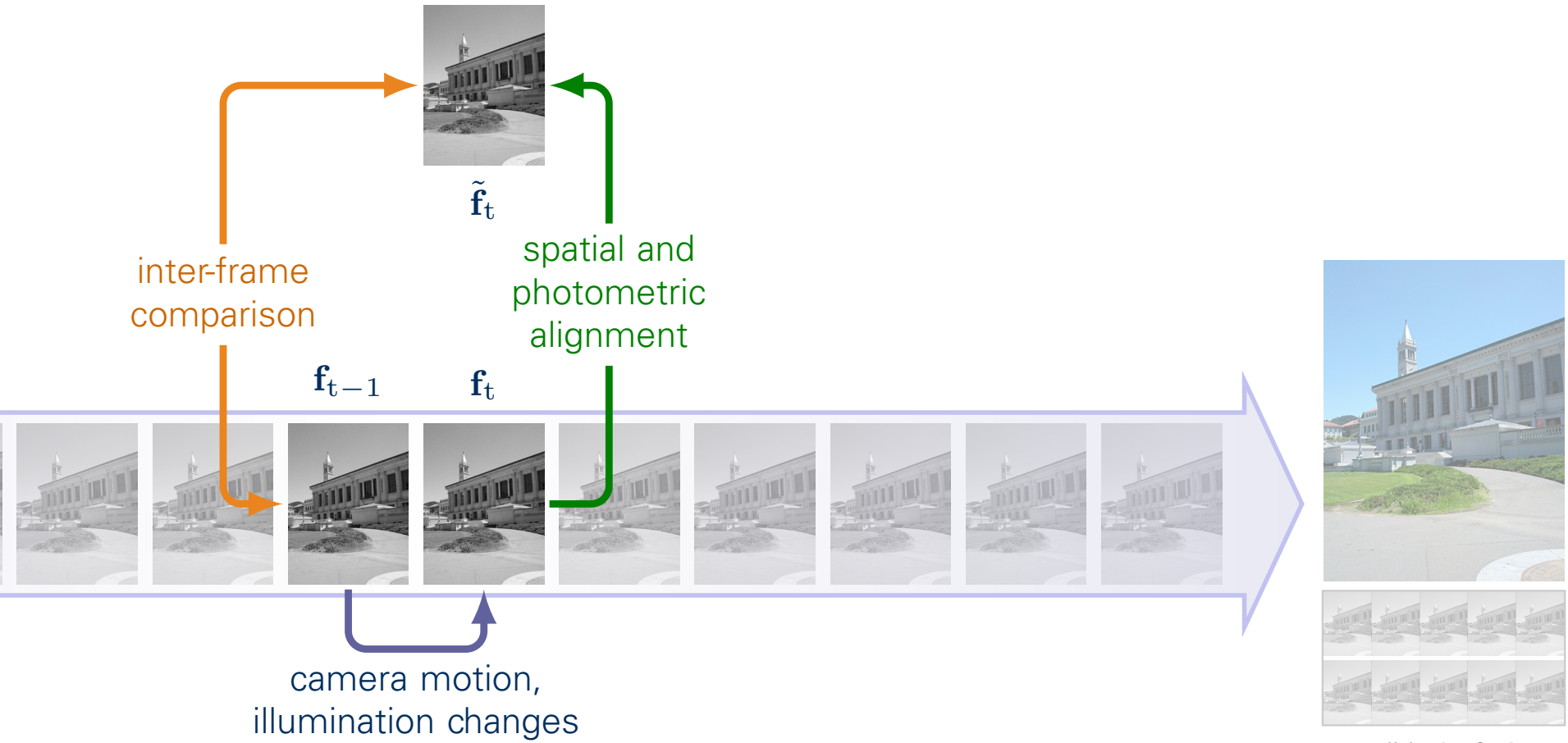
Frame alignment procedure

- record and store N preview frames prior to image capture ($\approx$10 fps, 320 $\times$ 240, JPG 85)



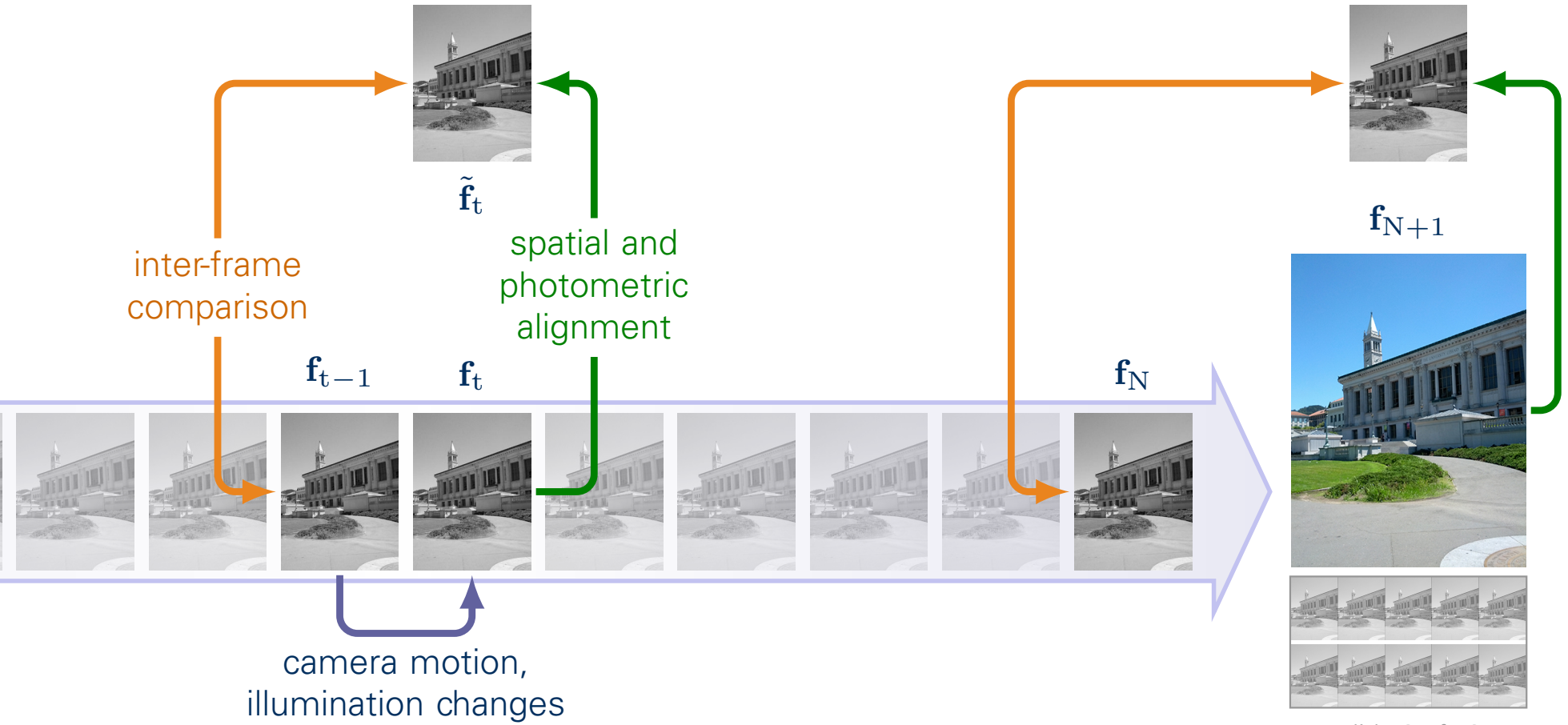
Frame alignment procedure

- record and store N preview frames prior to image capture ($\approx$10 fps, 320 $\times$ 240, JPG 85)
- pairwise alignment of consecutive frames to test for consistency



Frame alignment procedure

- record and store N preview frames prior to image capture ($\approx$10 fps, 320 $\times$ 240, JPG 85)
- pairwise alignment of consecutive frames to test for consistency
- full resolution image is treated as $N+1^{st}$ frame



Frame alignment in a nutshell

for each pair of consecutive frames $(\mathbf{f}_{t-1}, \mathbf{f}_t)$:

$\mathbf{f}_{N+1}$



$\mathbf{f}_N$



Frame alignment in a nutshell

for each pair of consecutive frames $(\mathbf{f}_{t-1}, \mathbf{f}_t)$:

- convert to grayscale, if necessary

$\mathbf{f}_{N+1}$



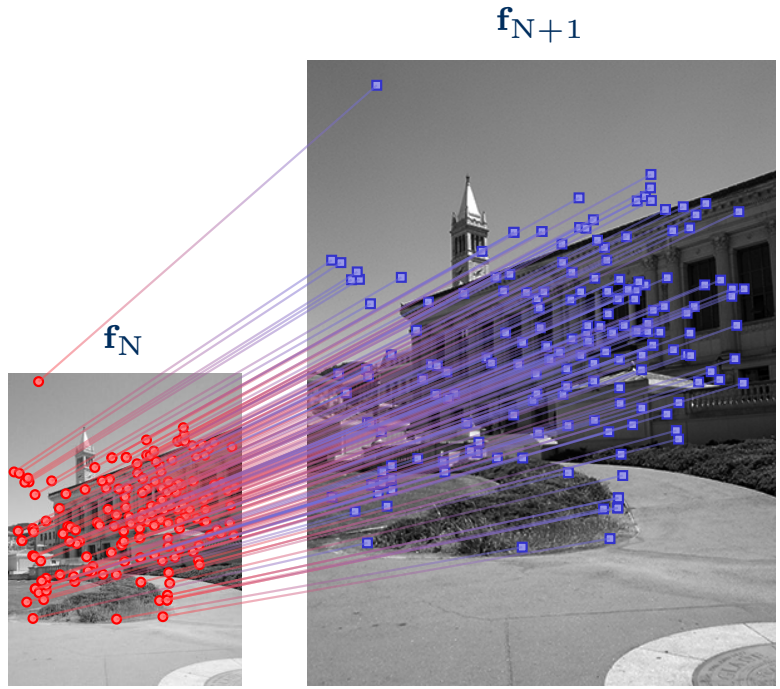
$\mathbf{f}_N$



Frame alignment in a nutshell

for each pair of consecutive frames $(\mathbf{f}_{t-1}, \mathbf{f}_t)$:

- convert to grayscale, if necessary
- find pairs of distinct SIFT key points $\{(\bigcirc, \square)\}$
  by matching key point descriptors $\bigcirc.\mathbf{d}$, $\square.\mathbf{d}$



D.G. Lowe, IJCV 2(60), 2004;

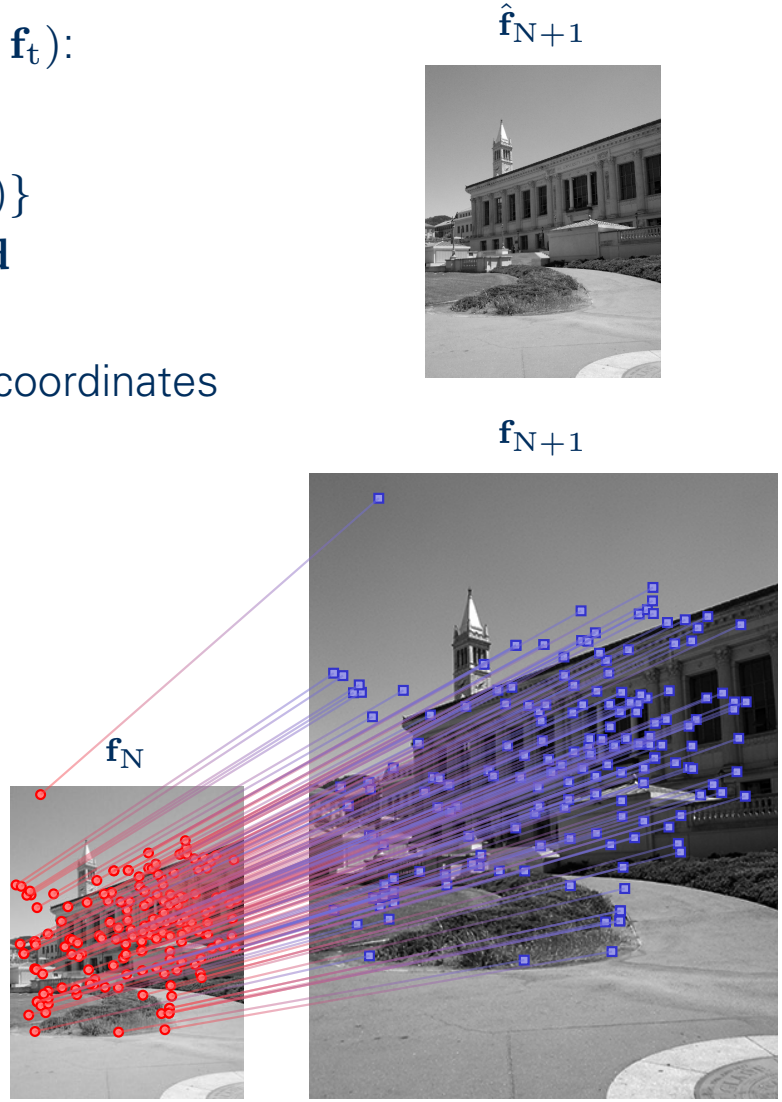
Frame alignment in a nutshell

for each pair of consecutive frames $(\mathbf{f}_{t-1}, \mathbf{f}_t)$:

- convert to grayscale, if necessary
- find pairs of distinct SIFT key points $\{(\bigcirc, \square)\}$
  by matching key point descriptors $\bigcirc.\mathbf{d}$, $\square.\mathbf{d}$
- use RANSAC to determine LMS-optimal
  global 3×3 homography $\mathbf{H}$ from key point coordinates

$$\begin{pmatrix} \bigcirc.x \\ \bigcirc.y \\ 1 \end{pmatrix} \approx \mathbf{H} \cdot \begin{pmatrix} \square.x \\ \square.y \\ 1 \end{pmatrix}$$

- warp $\mathbf{f}_t$ according to $\mathbf{H}$ to yield $\hat{\mathbf{f}}_t$



D.G. Lowe, IJCV 2(60), 2004;
M.A. Fischler et. al, CACM 6(24), 1981;

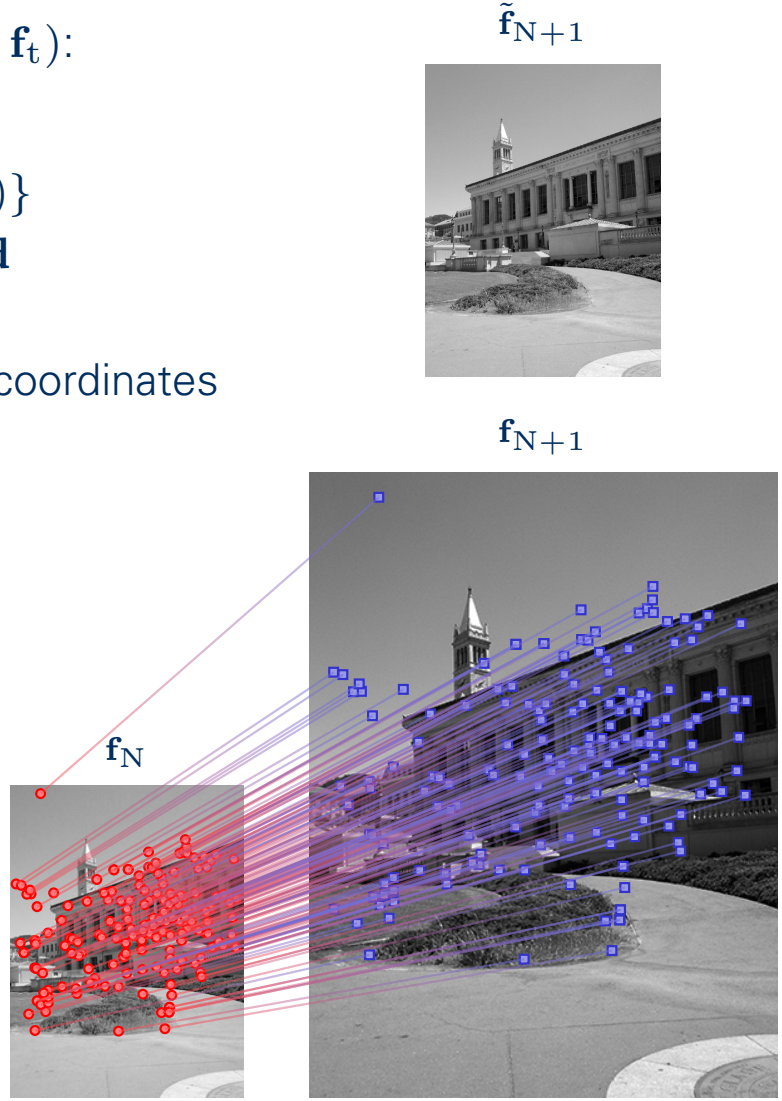
Frame alignment in a nutshell

for each pair of consecutive frames $(\mathbf{f}_{t-1}, \mathbf{f}_t)$:

- convert to grayscale, if necessary
- find pairs of distinct SIFT key points $\{(\bigcirc, \square)\}$ by matching key point descriptors $\bigcirc.\mathbf{d}$, $\square.\mathbf{d}$
- use RANSAC to determine LMS-optimal global 3×3 homography $\mathbf{H}$ from key point coordinates

$$\begin{pmatrix} \bigcirc.x \\ \bigcirc.y \\ 1 \end{pmatrix} \approx \mathbf{H} \cdot \begin{pmatrix} \square.x \\ \square.y \\ 1 \end{pmatrix}$$

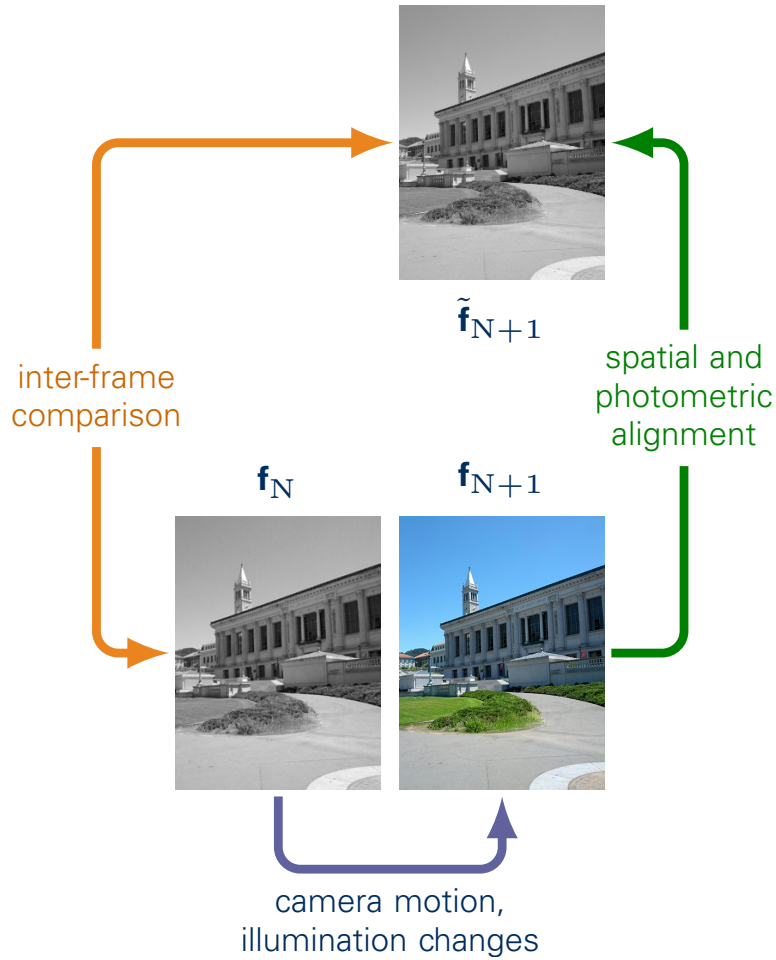
- warp $\mathbf{f}_t$ according to $\mathbf{H}$ to yield $\hat{\mathbf{f}}_t$
- match histograms to yield photometrically aligned frames $(\mathbf{f}_{t-1}, \tilde{\mathbf{f}}_t)$ with identical histograms



D.G. Lowe, IJCV 2(60), 2004;
M.A. Fischler et. al, CACM 6(24), 1981;
J. Morovic et al., Pattern Recogn Lett 23, 2002

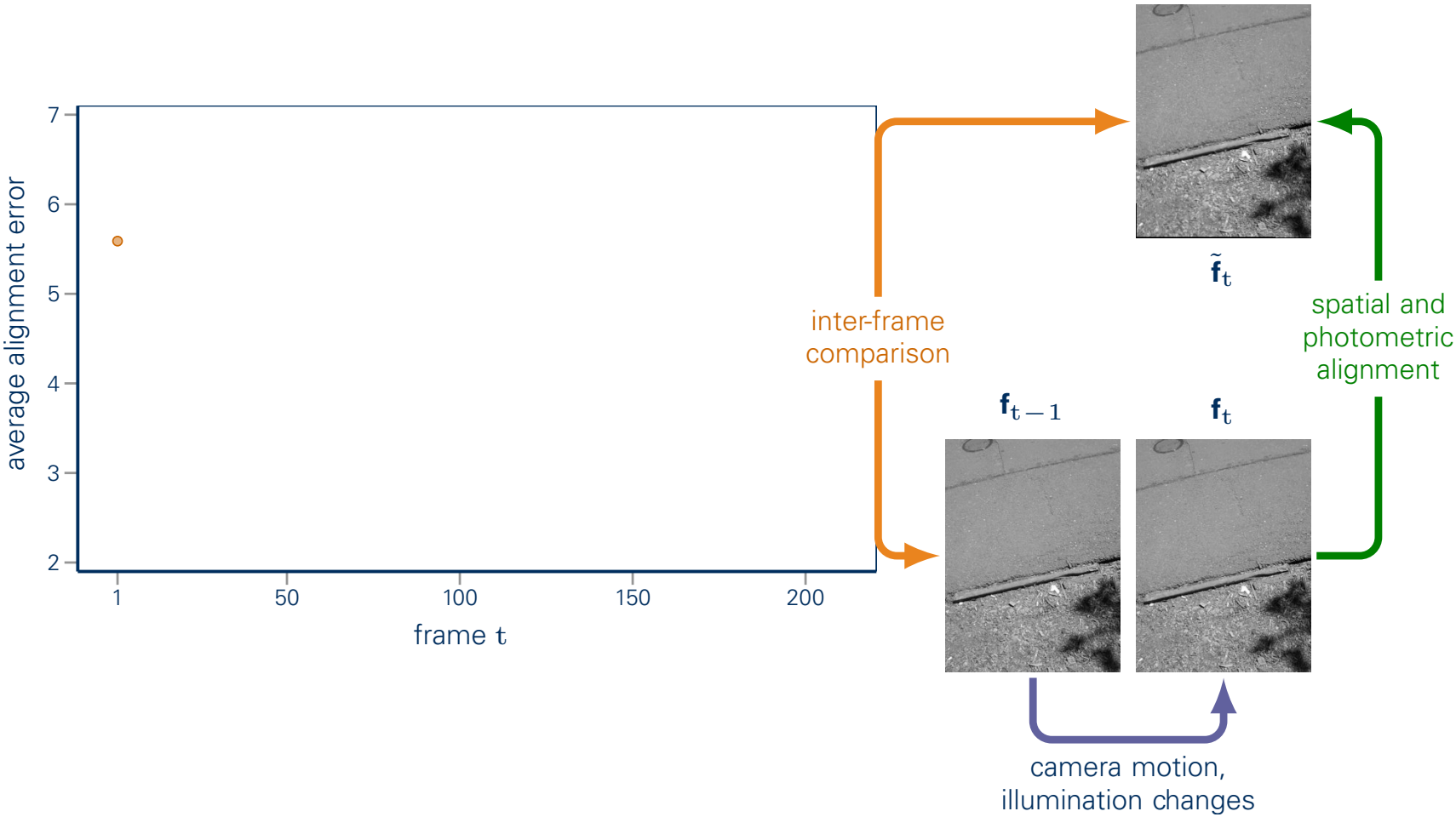
Frame alignment: Example

- compare aligned frames in terms of pixel-wise absolute alignment error $|\tilde{\mathbf{f}}_t - \mathbf{f}_{t-1}|$



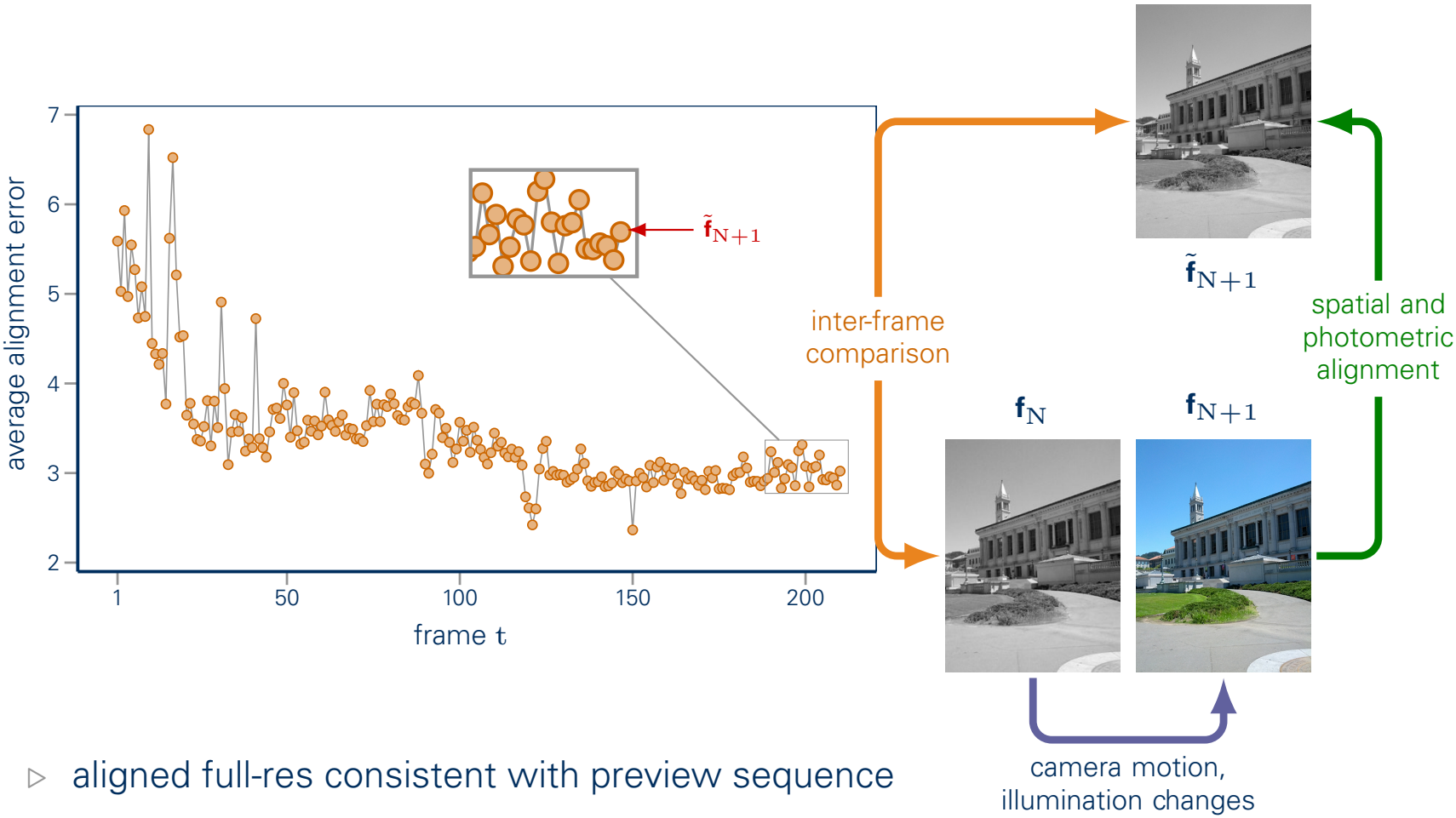
Frame alignment: Example

- compare aligned frames in terms of pixel-wise absolute alignment error $|\tilde{\mathbf{f}}_t - \mathbf{f}_{t-1}|$



Frame alignment: Example

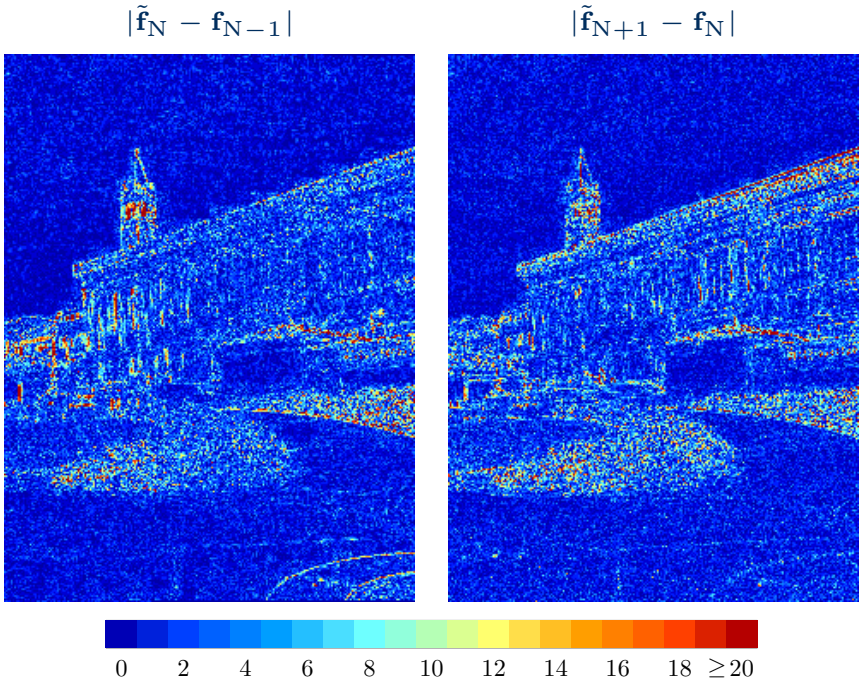
- compare aligned frames in terms of pixel-wise absolute alignment error $|\tilde{\mathbf{f}}_t - \mathbf{f}_{t-1}|$

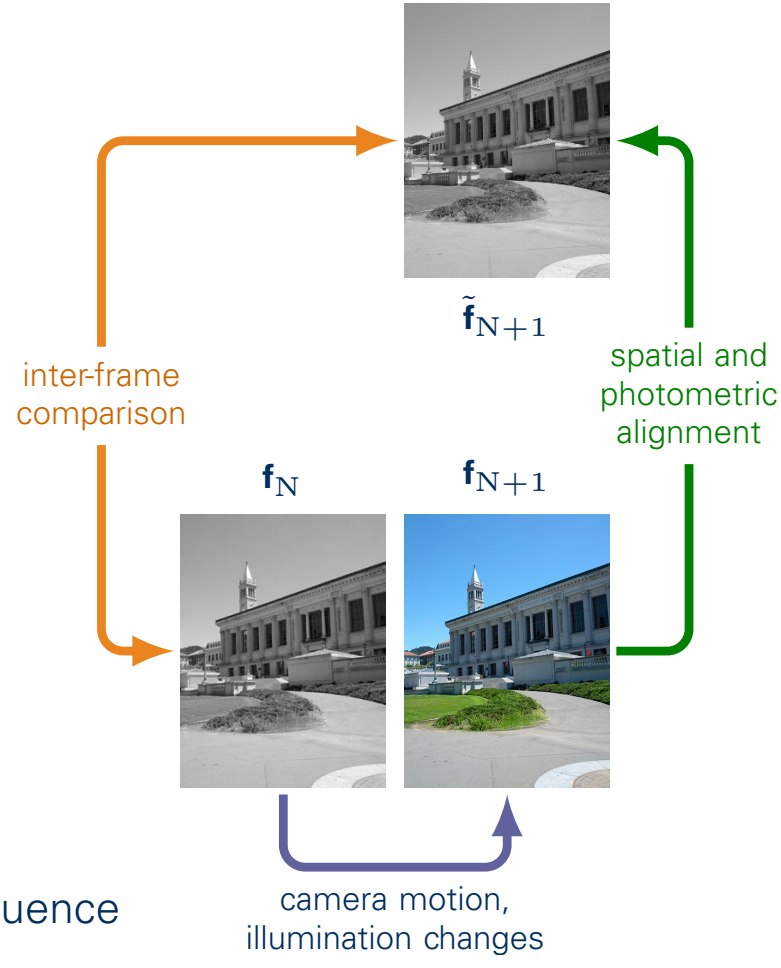


- ▷ aligned full-res consistent with preview sequence

Alignment error

- compare aligned frames in terms of pixel-wise absolute alignment error $|\tilde{\mathbf{f}}_t - \mathbf{f}_{t-1}|$

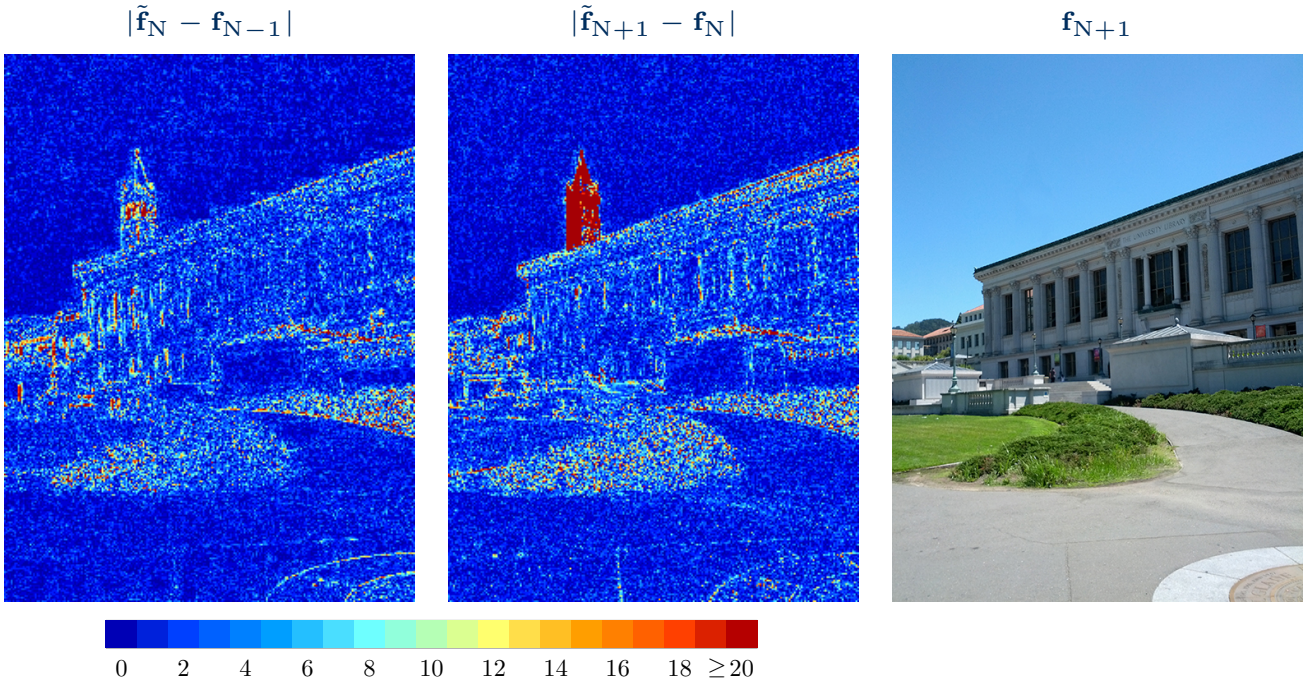




- ▷ aligned full-res consistent with preview sequence

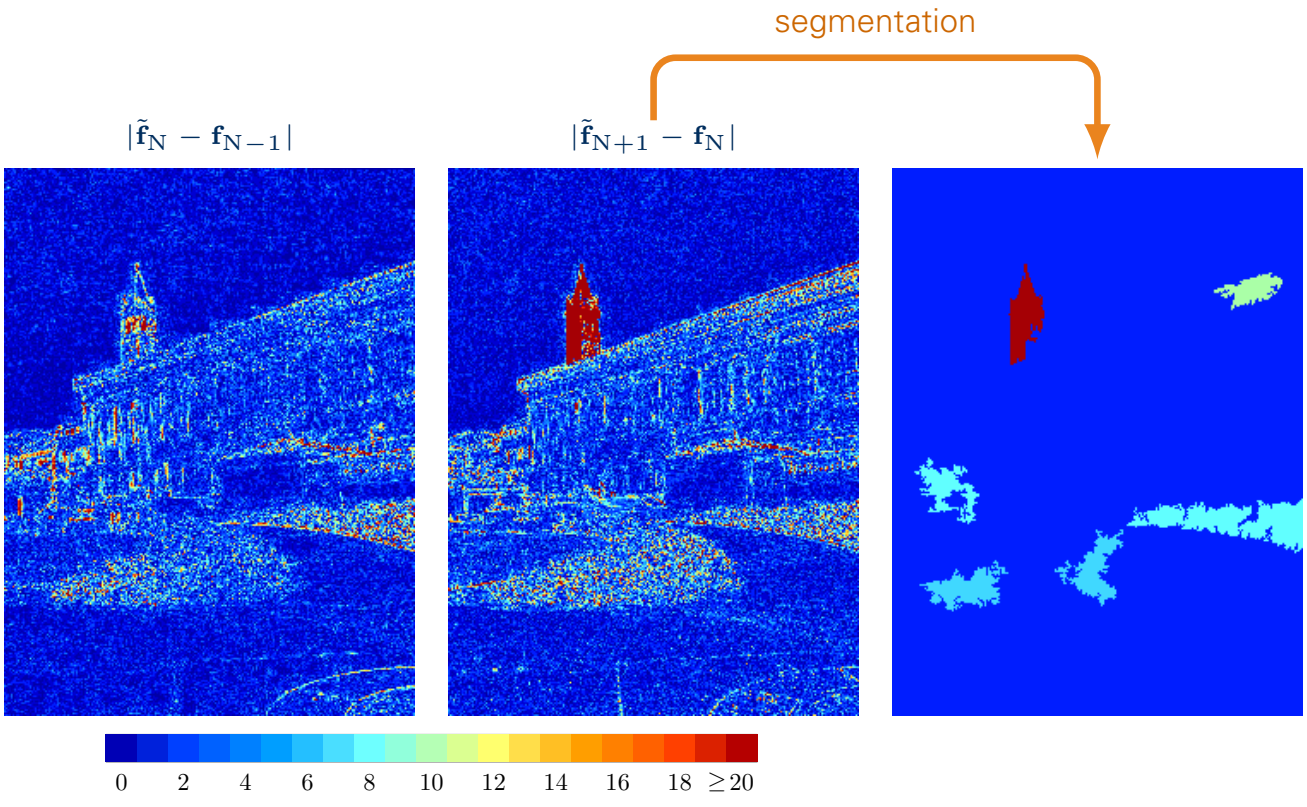
Alignment error: Manipulation

- manipulated image regions yield increased alignment error $|\tilde{\mathbf{f}}_{\mathrm{N}+1} - \mathbf{f}_{\mathrm{N}}|$



  - ▷ aligned full-res inconsistent with preview sequence

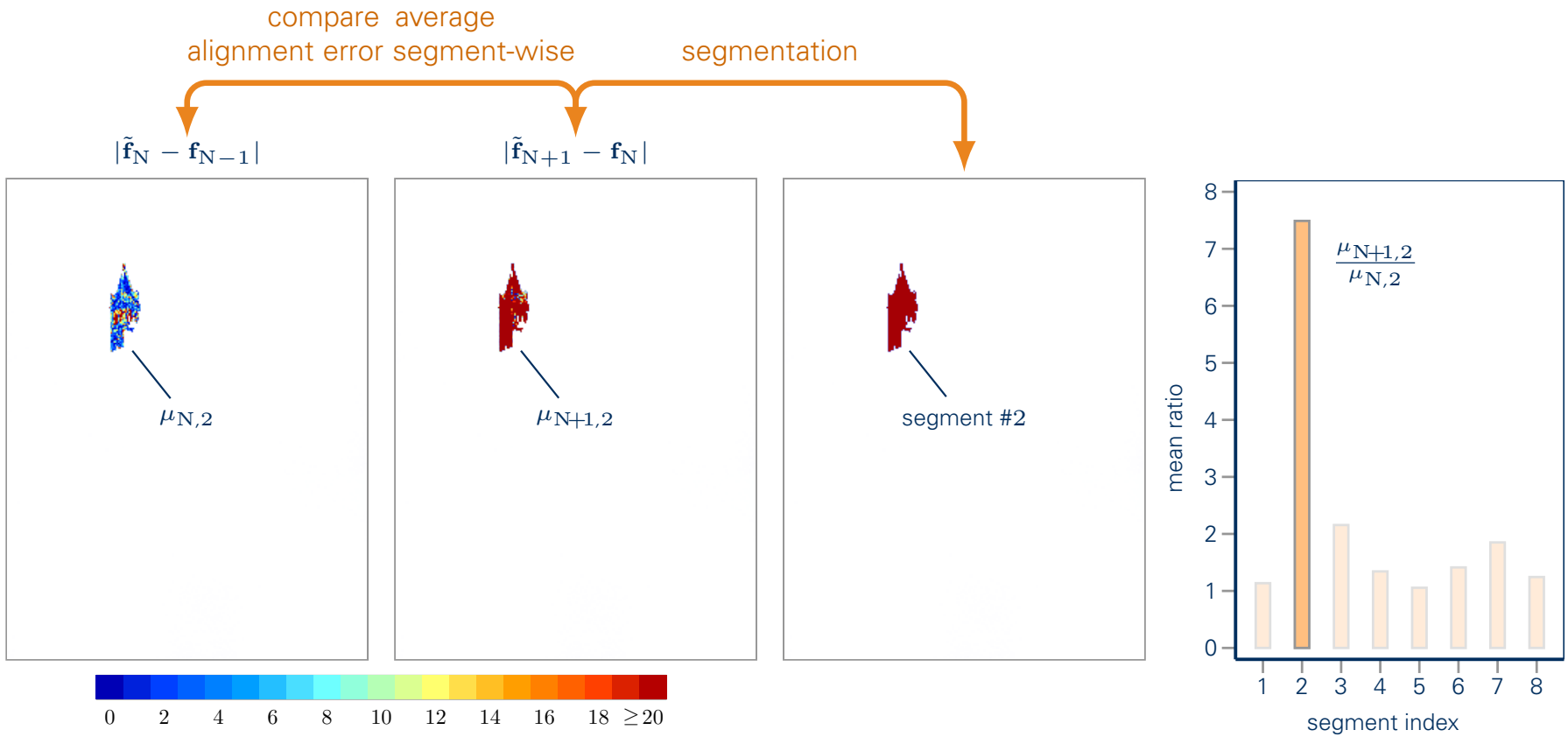
Automatic detection procedure



- coarse mean-shift segmentation of the alignment error

D. Comaniciu et al., PAMI 24(5), 2002

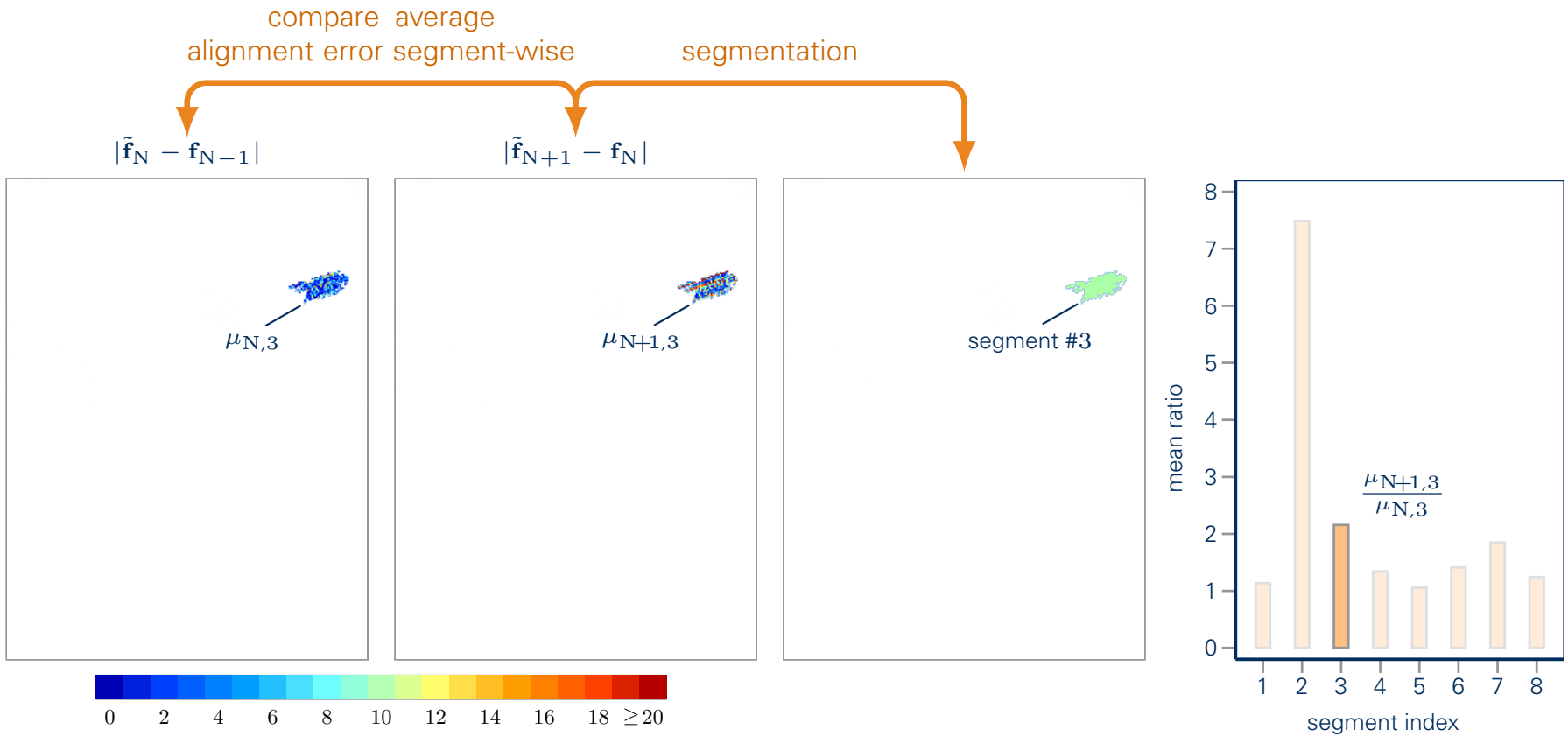
Automatic detection procedure



- coarse mean-shift segmentation of the alignment error

D. Comaniciu et al., PAMI 24(5), 2002

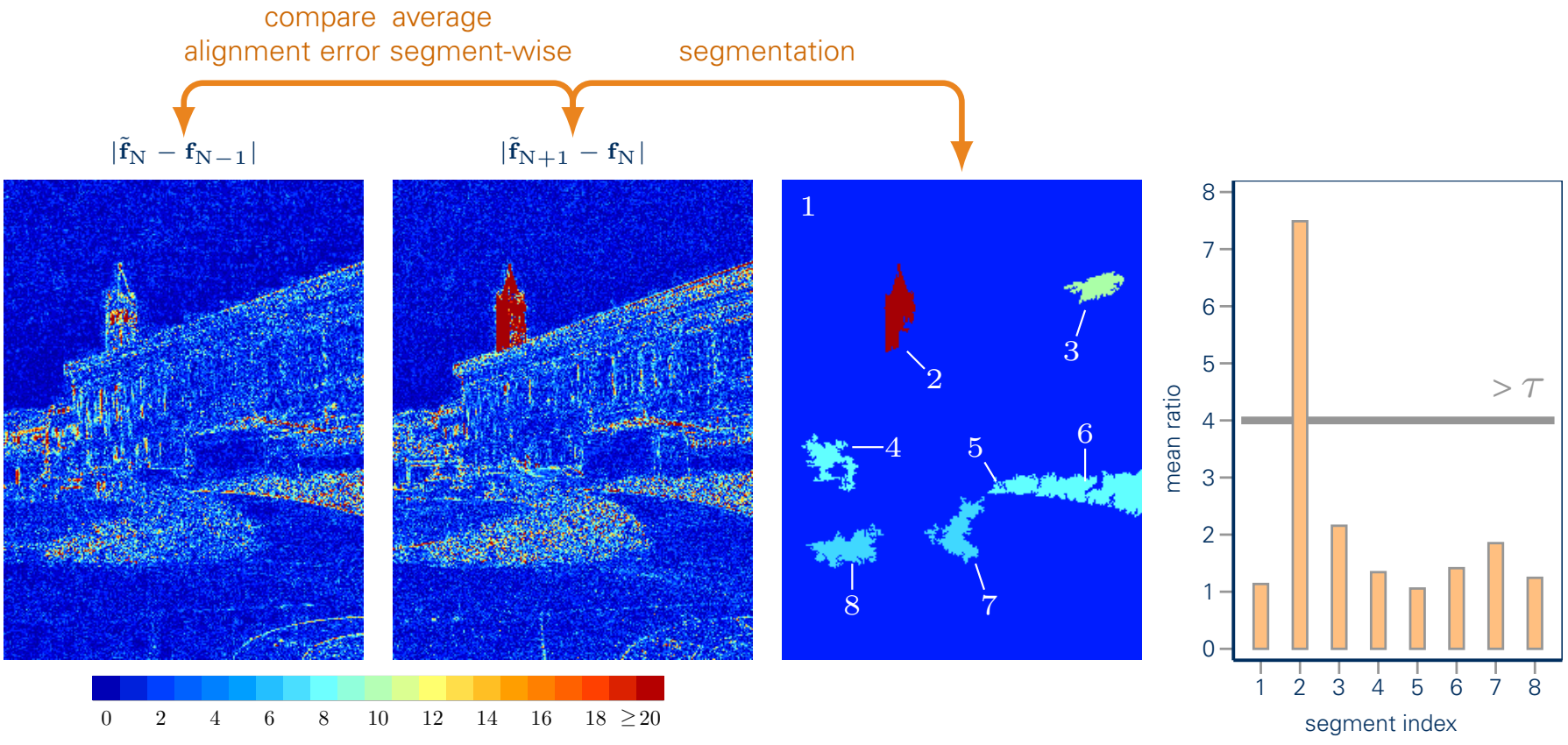
Automatic detection procedure



- coarse mean-shift segmentation of the alignment error

D. Comaniciu et al., PAMI 24(5), 2002

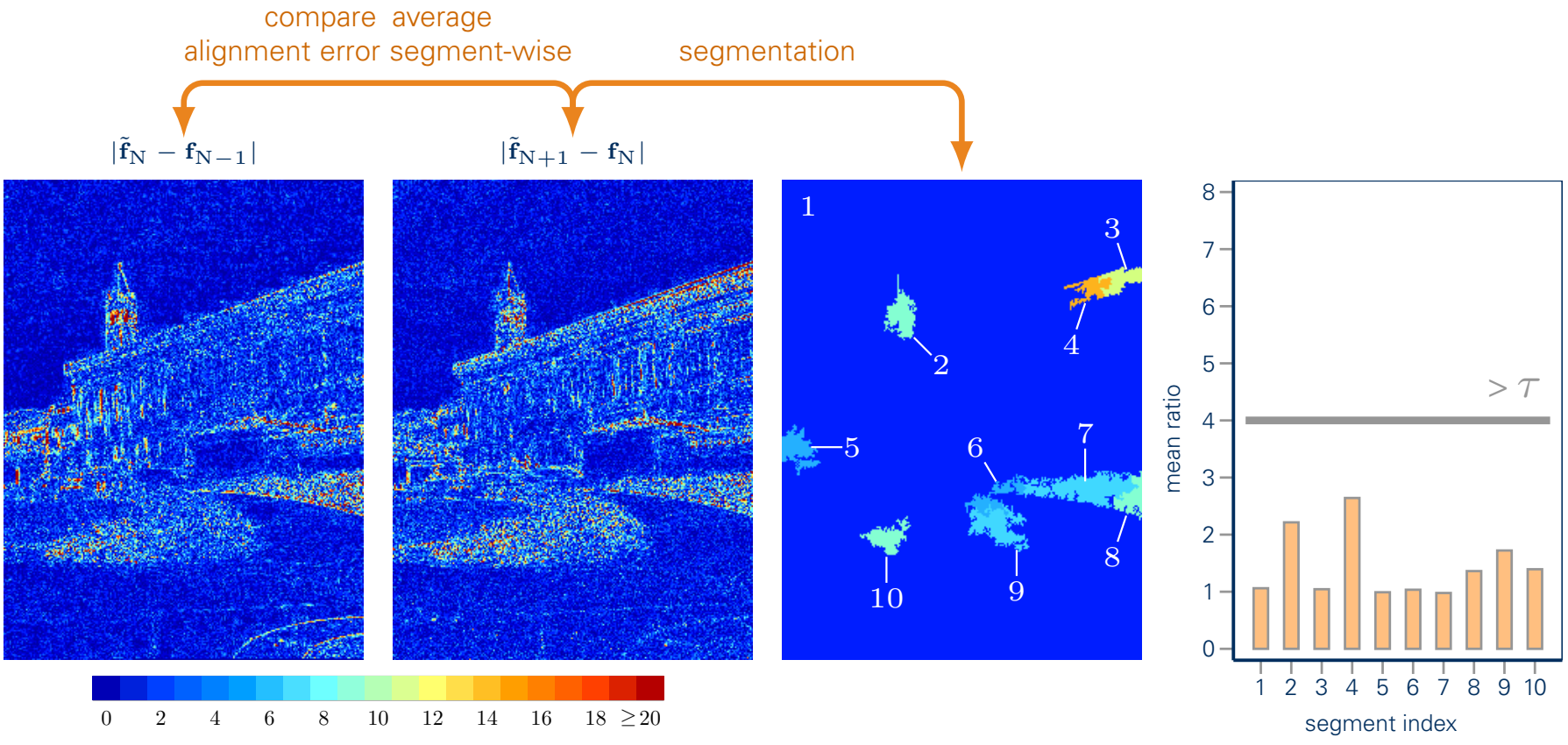
Automatic detection procedure



- coarse mean-shift segmentation of the alignment error
- regions with relative error larger than a specified threshold τ are labeled as potentially altered

D. Comaniciu et al., PAMI 24(5), 2002

Automatic detection procedure

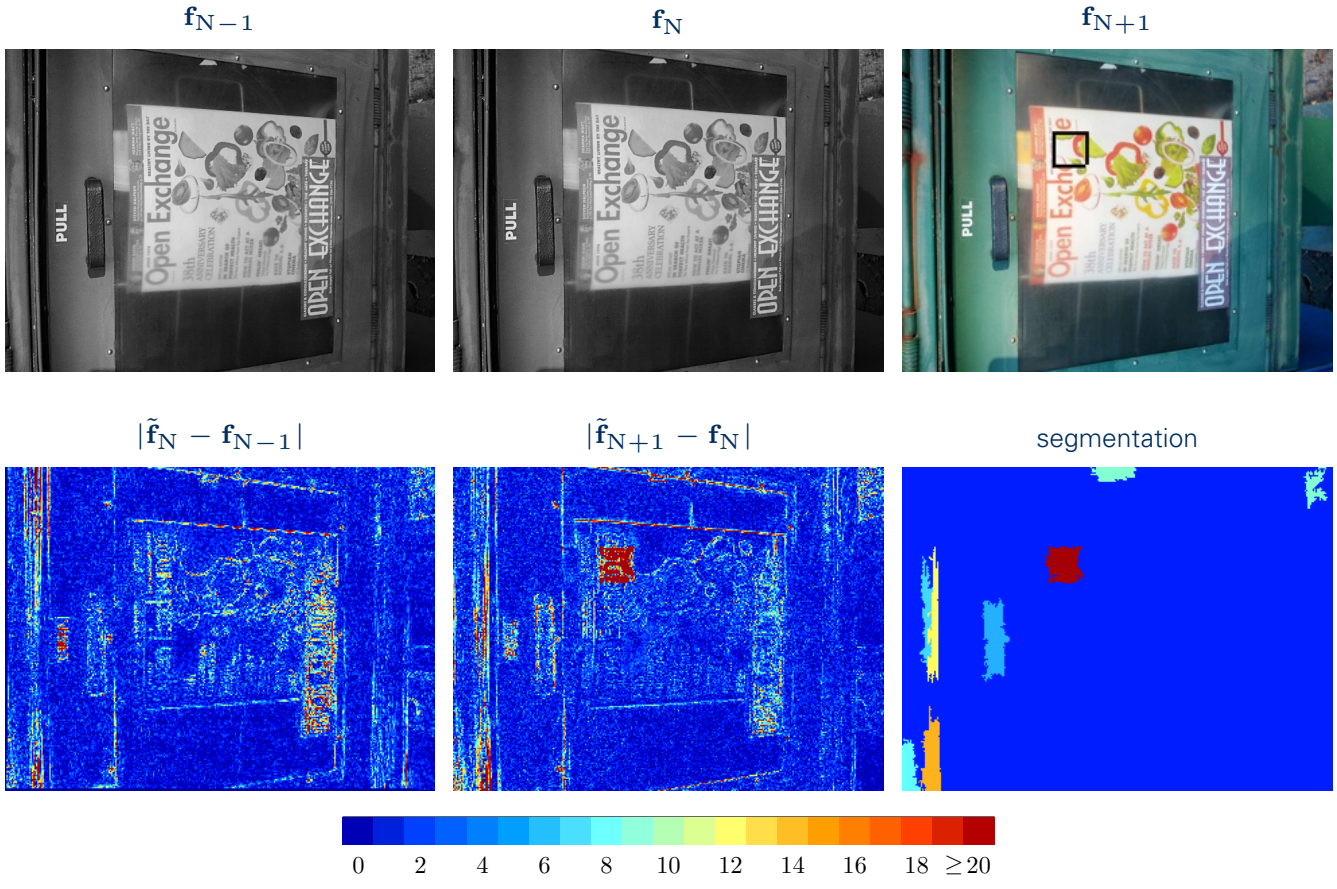


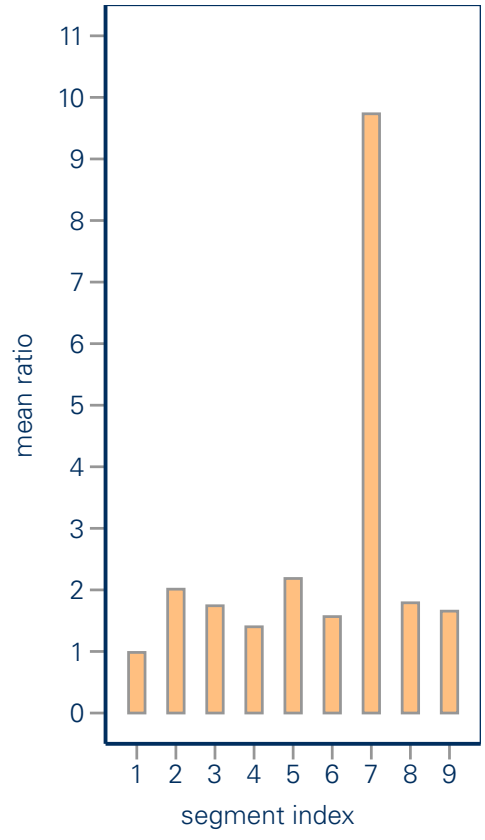
- coarse mean-shift segmentation of the alignment error
- regions with relative error larger than a specified threshold τ are labeled as potentially altered

D. Comaniciu et al., PAMI 24(5), 2002

More examples

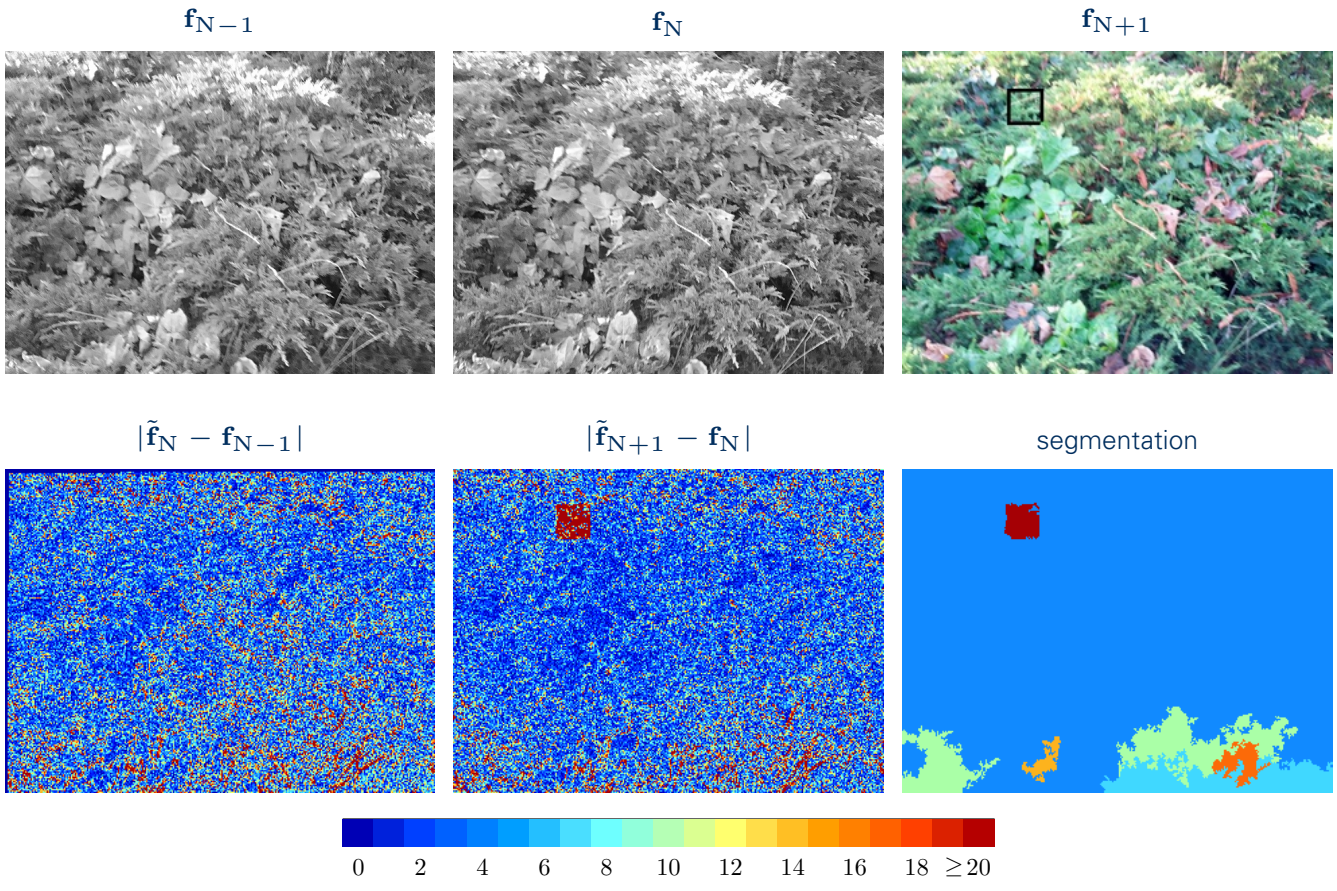
- duplication of a randomly chosen 200×200 image region
- original image size 1944×2592; preview size 240×320, saved at JPG 85

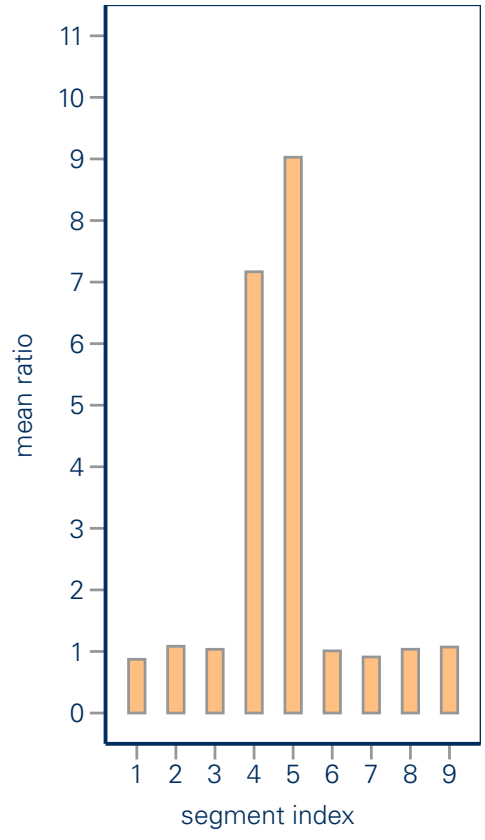




More examples

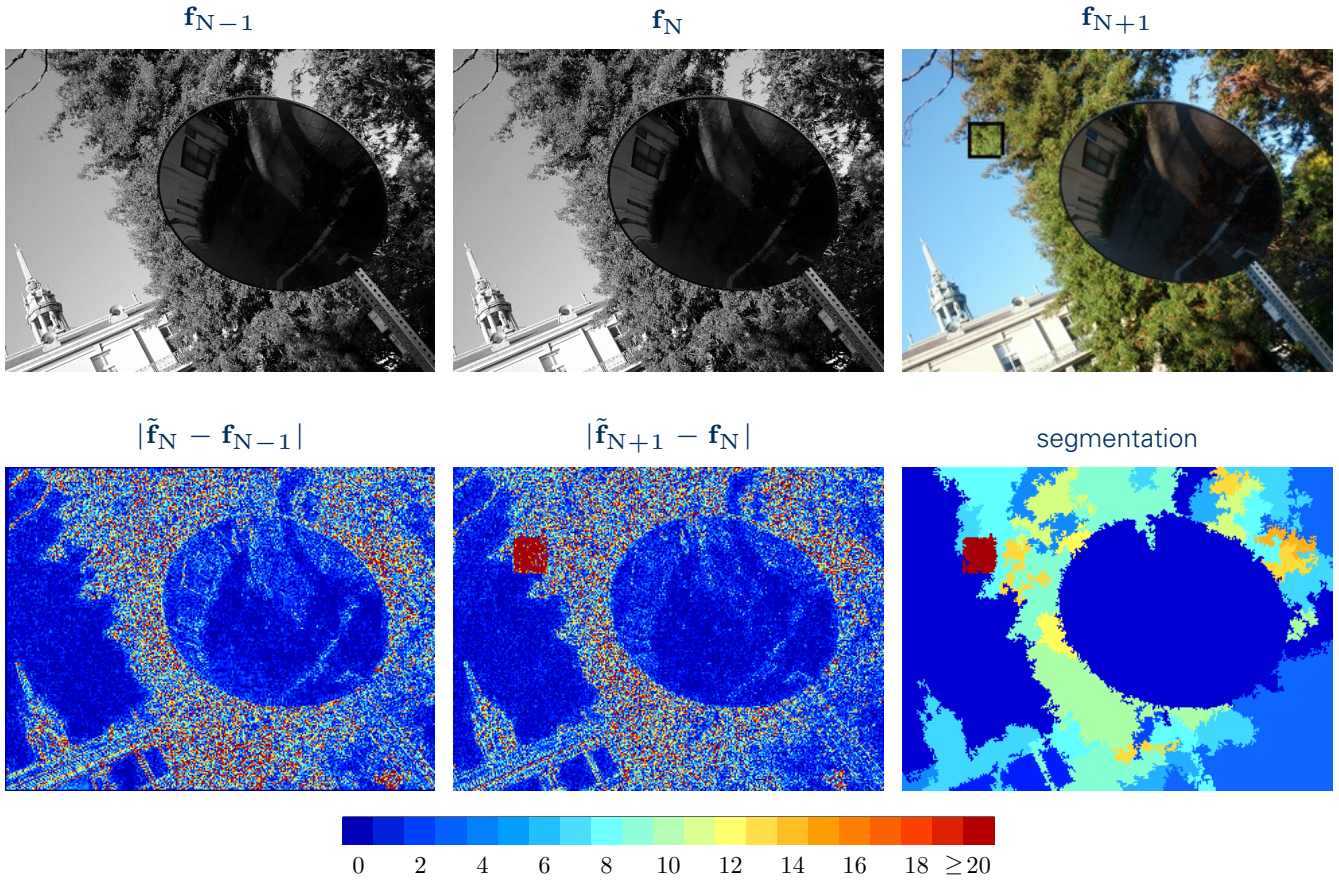
- duplication of a randomly chosen 200×200 image region
- original image size 1944×2592; preview size 240×320, saved at JPG 85

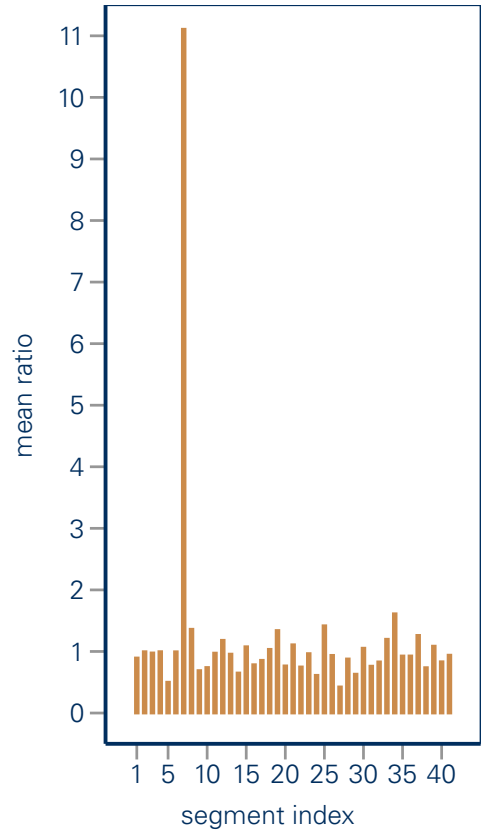




More examples

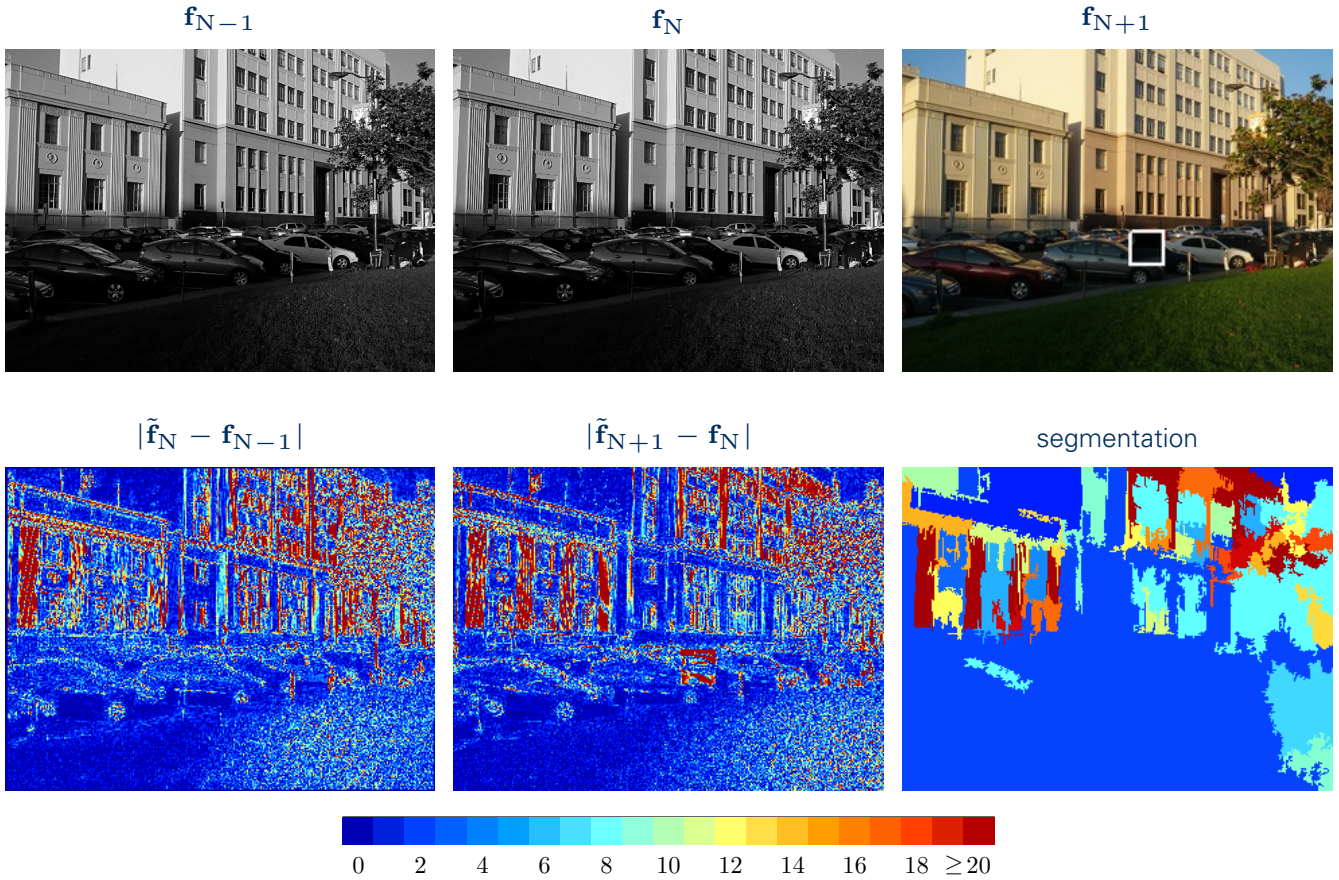
- duplication of a randomly chosen 200×200 image region
- original image size 1944×2592; preview size 240×320, saved at JPG 85

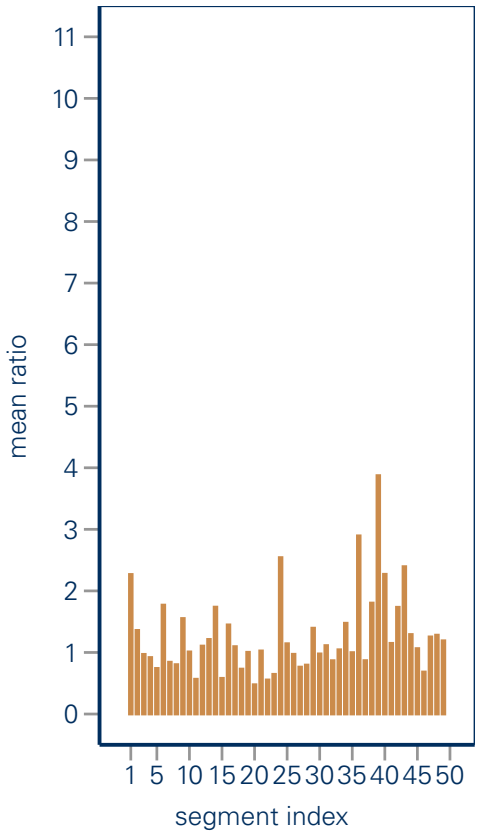




More examples

- duplication of a randomly chosen 200×200 image region
- original image size 1944×2592; preview size 240×320, saved at JPG 85

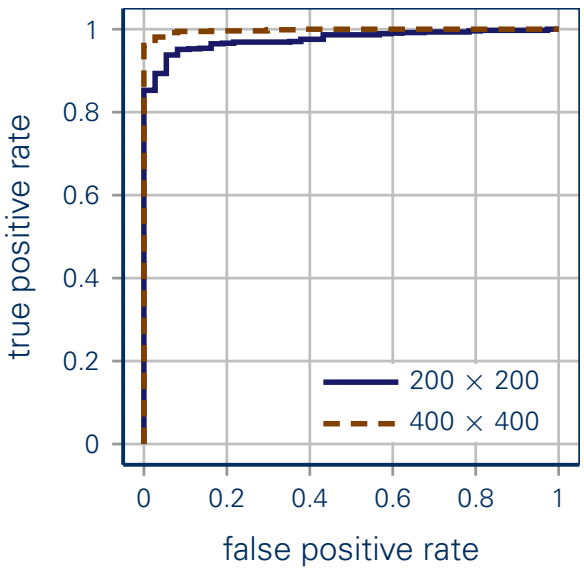




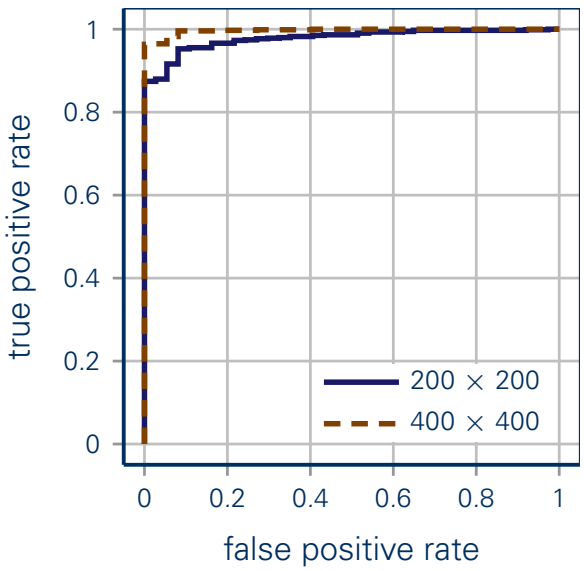
Quantitative results

- 20 random duplications from each 37 original images (size: 1944×2592)
- decision criterion: maximum mean ratio

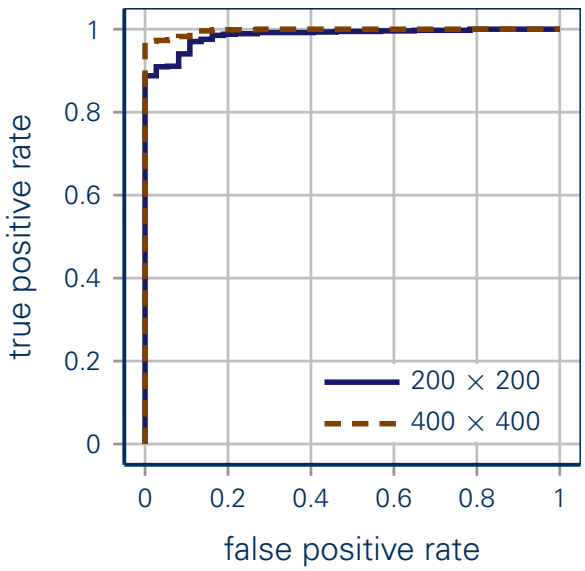
240 × 320, JPG 85



240 × 320, JPG 95

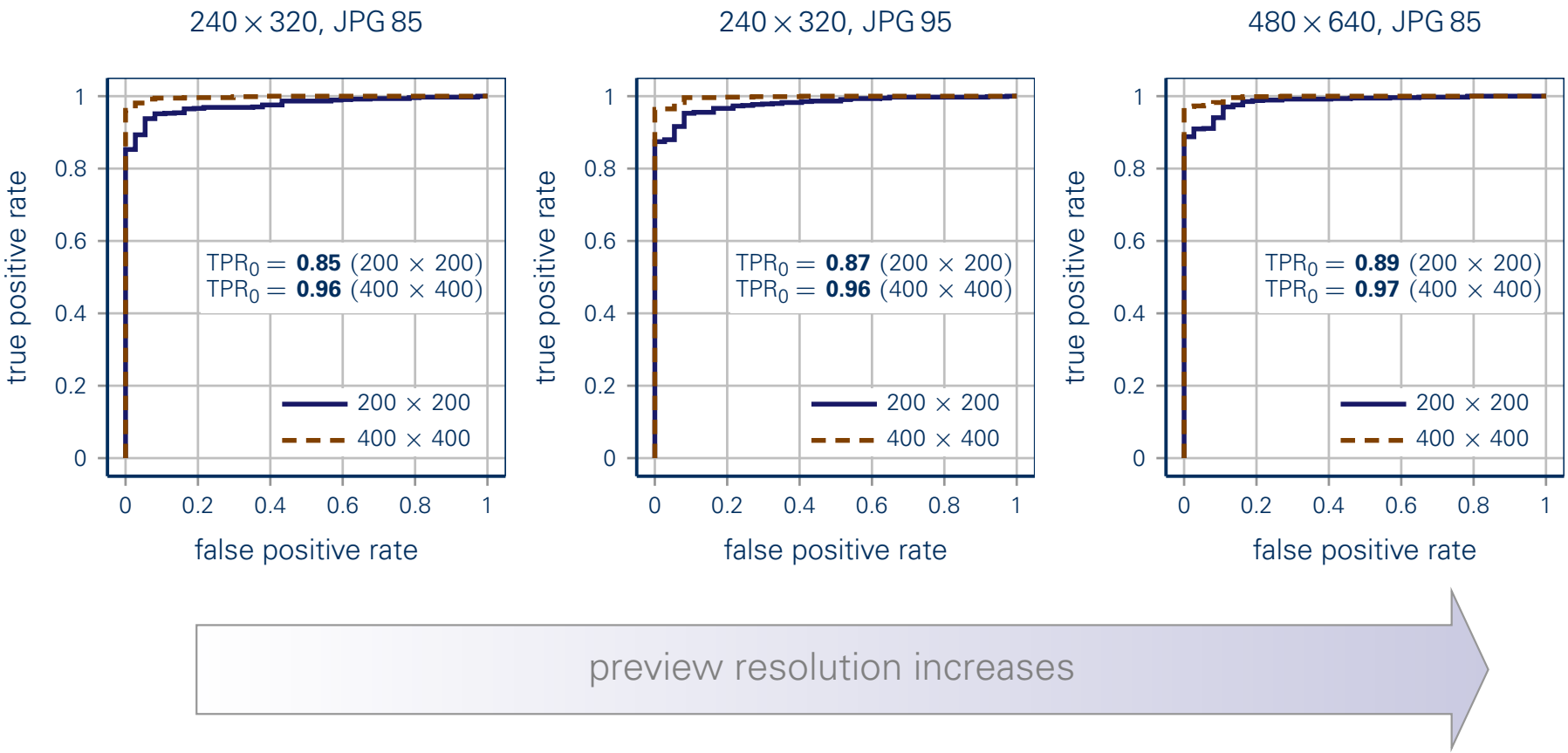


480 × 640, JPG 85

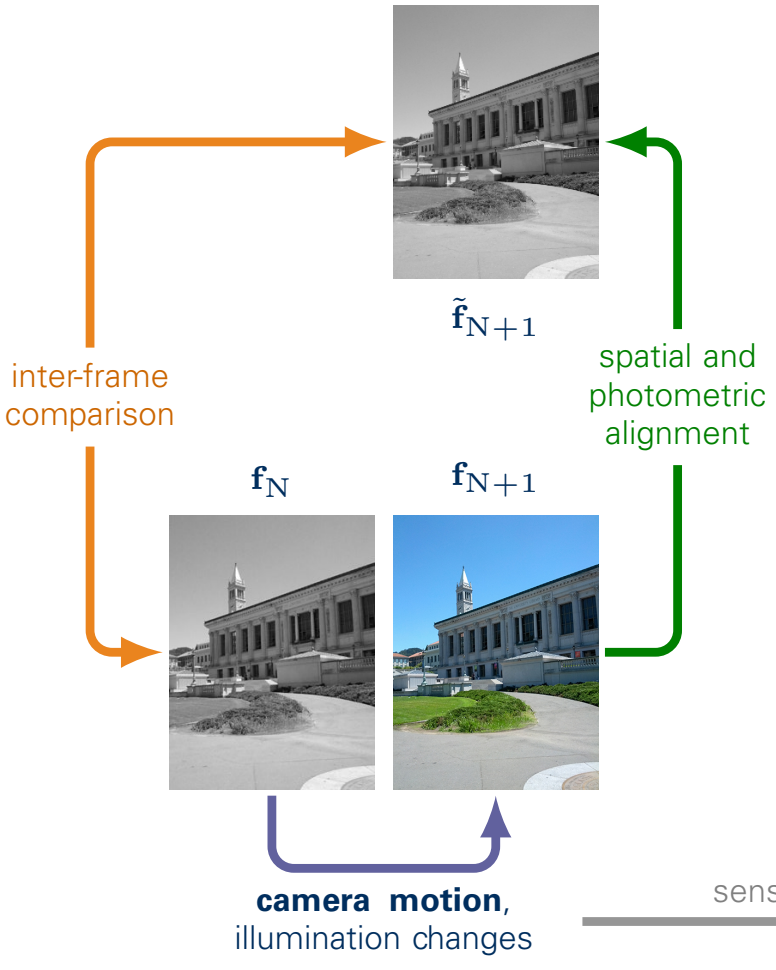


Quantitative results

- 20 random duplications from each 37 original images (size: 1944×2592)
- decision criterion: maximum mean ratio

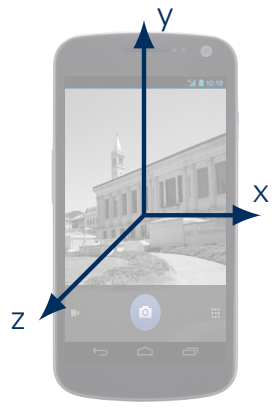


Camera motion

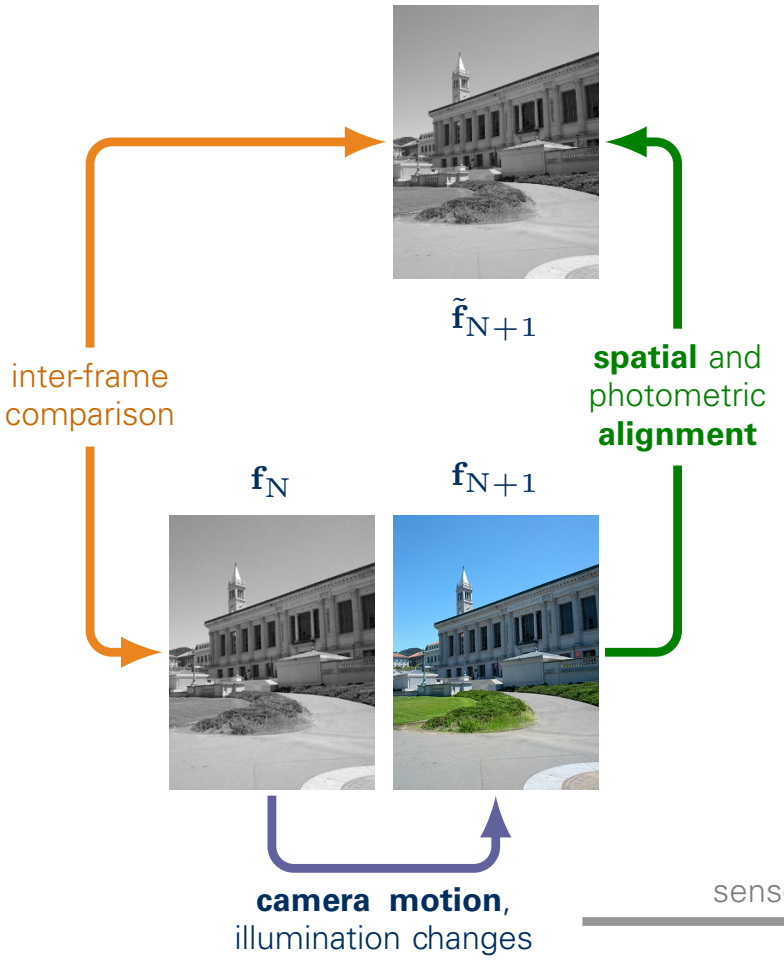


- typical smartphones are equipped with motion sensors, e. g., gyroscope and accelerometer

sensor → rotation and acceleration along x,y, and z axes



Camera motion

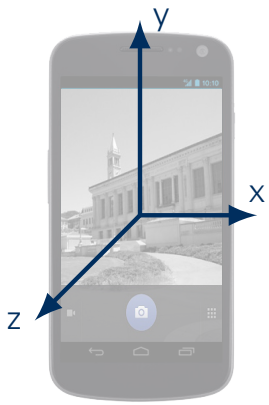


- typical smartphones are equipped with motion sensors, e. g., gyroscope and accelerometer
- if stored with the image, plausible forgeries need to be consistent with recorded sensor measurements

scene → homography $\mathbf{H}_t$

comparison between scene-based and sensor-based motion estimates

sensor → rotation and acceleration along x,y, and z axes



Camera motion measurements

image-based

- deviation of the inter-frame alignment $\mathbf{H}$ from the identity matrix $\mathbf{I}$

$$\|\mathbf{H} - \mathbf{I}\|_{\mathrm{F}} = \sum_{\mathrm{i,j}} \left(\mathrm{H_{i,j}} - \mathrm{I_{i,j}}\right)^2$$

sensor-based

- absolute rotational angle averaged over all three axes

Camera motion measurements

image-based

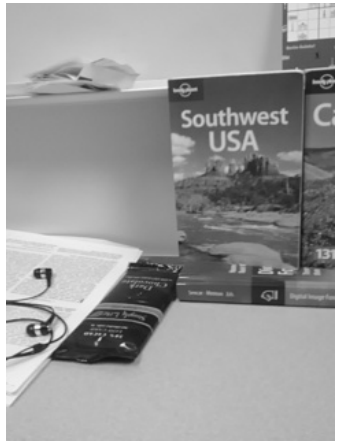
- deviation of the inter-frame alignment $\mathbf{H}$ from the identity matrix $\mathbf{I}$

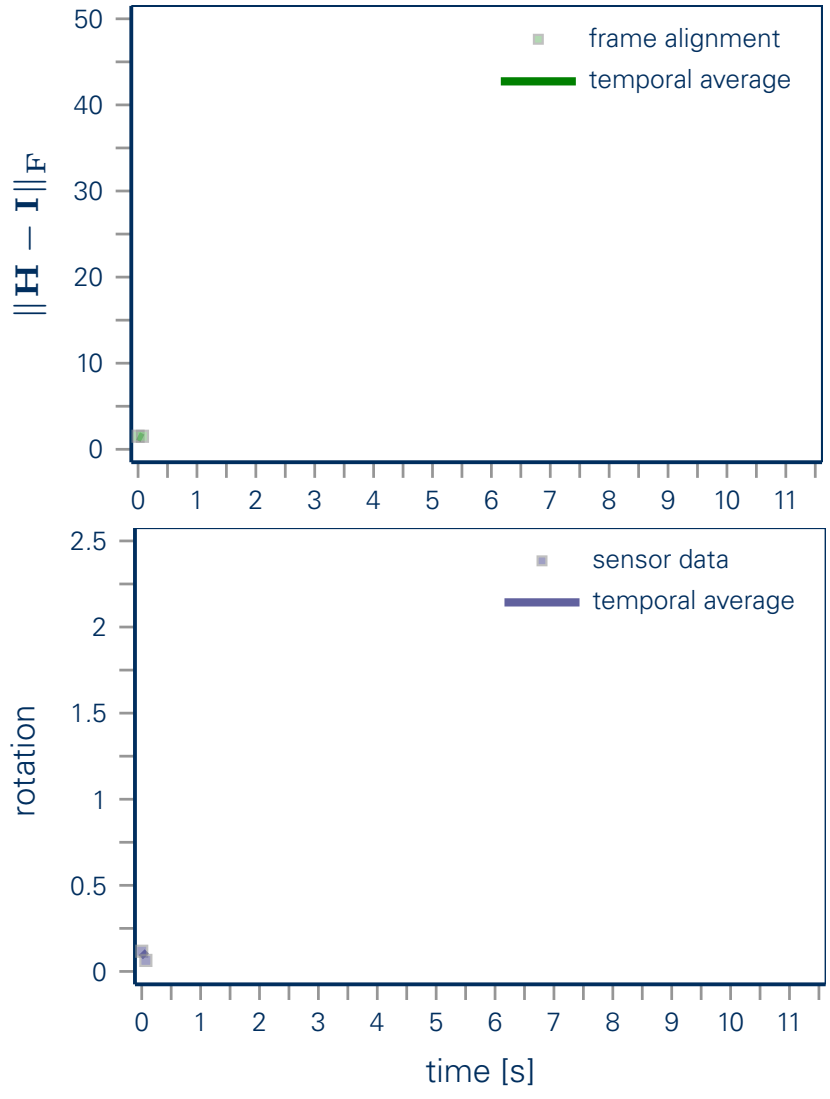
$$\|\mathbf{H} - \mathbf{I}\|_{\mathrm{F}} = \sum_{\mathrm{i,j}} \left(\mathrm{H}_{\mathrm{i,j}} - \mathrm{I}_{\mathrm{i,j}}\right)^2$$

sensor-based

- absolute rotational angle averaged over all three axes

preview sequence ≈10fps
gyroscope ≈12fps





Camera motion measurements

image-based

- deviation of the inter-frame alignment $\mathbf{H}$ from the identity matrix $\mathbf{I}$

$$\|\mathbf{H}-\mathbf{I}\|_{\mathrm{F}}=\sum_{\mathrm{i,j}}\left(\mathrm{H}_{\mathrm{i,j}}-\mathrm{I}_{\mathrm{i,j}}\right)^2$$

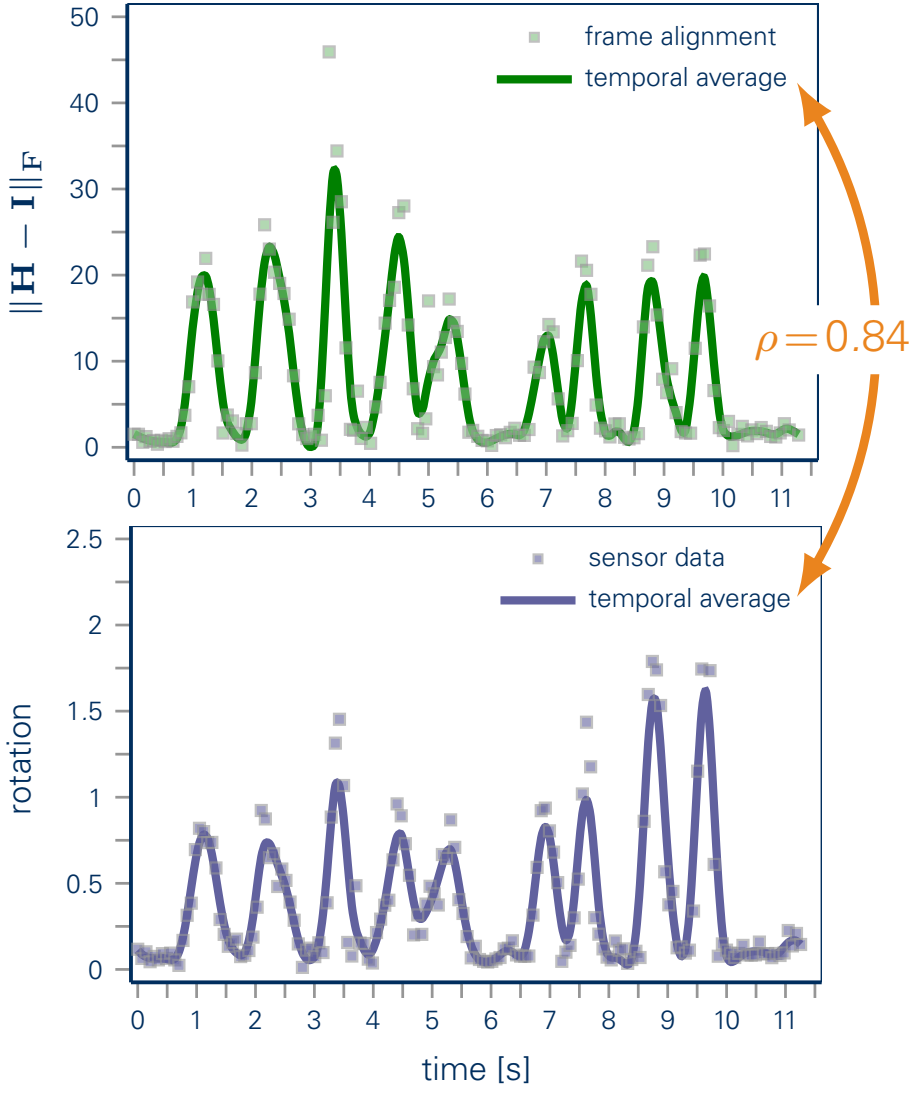
sensor-based

- absolute rotational angle averaged over all three axes

preview sequence ≈10fps

gyroscope ≈12fps





Camera motion measurements

image-based

- deviation of the inter-frame alignment $\mathbf{H}$ from the identity matrix $\mathbf{I}$

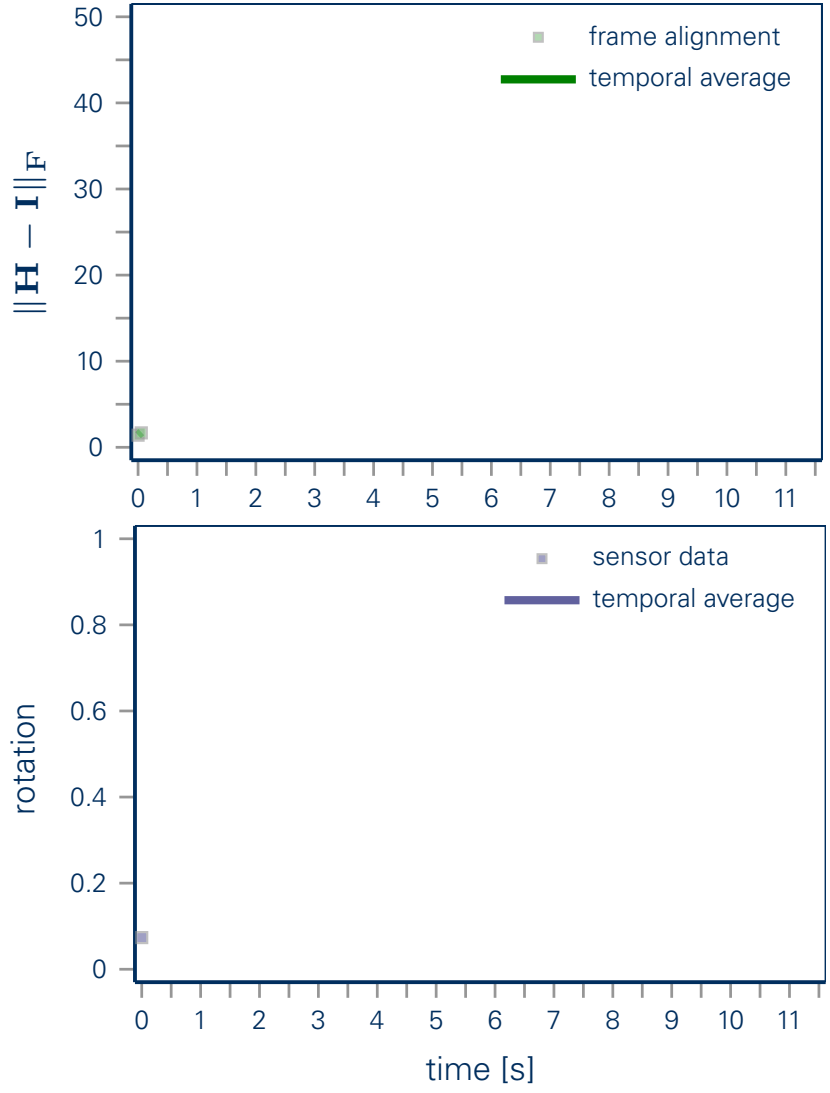
$$\|\mathbf{H} - \mathbf{I}\|_{\mathrm{F}} = \sum_{\mathrm{i,j}} \left(\mathrm{H_{i,j}} - \mathrm{I_{i,j}}\right)^2$$

sensor-based

- absolute rotational angle averaged over all three axes

preview sequence ≈10fps
gyroscope ≈12fps





Camera motion measurements

image-based

- deviation of the inter-frame alignment $\mathbf{H}$ from the identity matrix $\mathbf{I}$

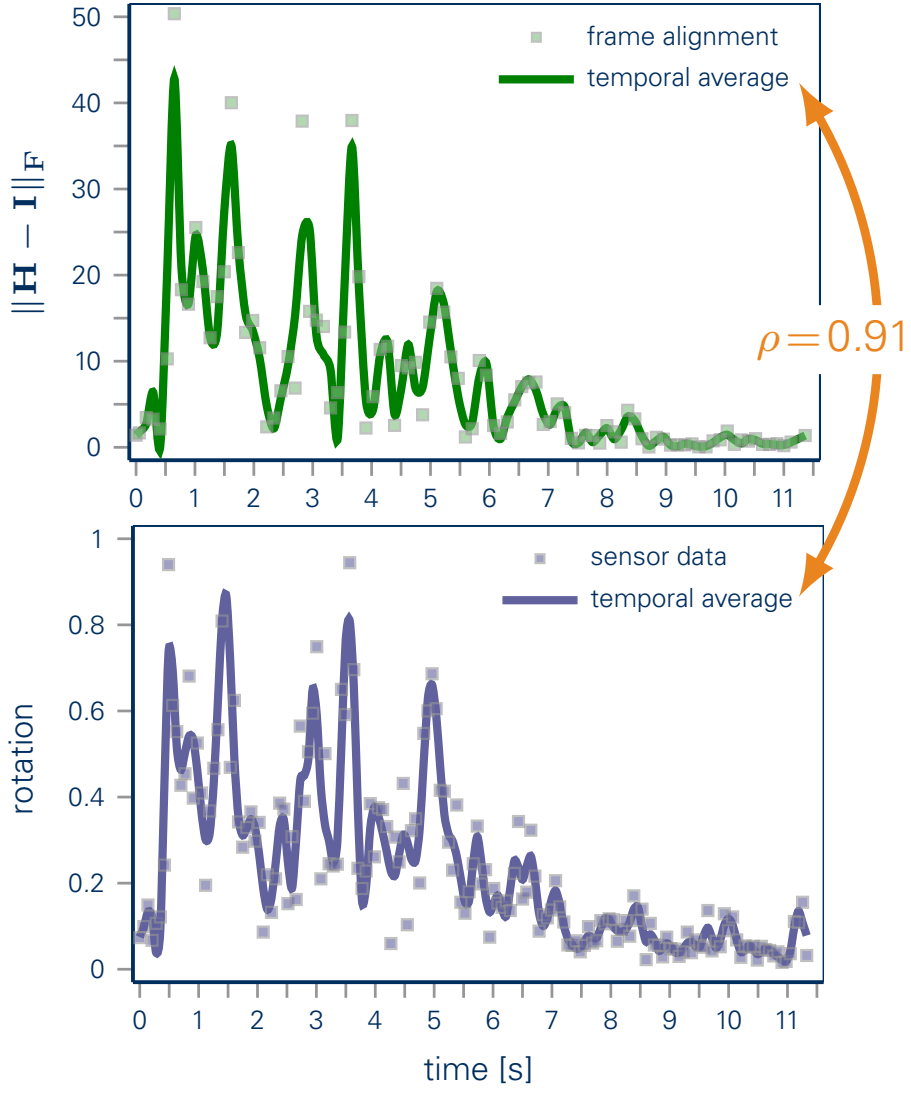
$$\|\mathbf{H}-\mathbf{I}\|_{\mathrm{F}}=\sum_{\mathrm{i,j}}\left(\mathrm{H}_{\mathrm{i,j}}-\mathrm{I}_{\mathrm{i,j}}\right)^2$$

sensor-based

- absolute rotational angle averaged over all three axes

preview sequence ≈10fps
gyroscope ≈12fps





Summary

- the more information that a camera records, the more difficult and time consuming it will be to create a compelling fake

sequence of preview frames

- manipulation needs to be propagated through all recorded frames

motion sensor data

- (forged) preview needs to be consistent with the recorded camera motion

Summary

- the more information that a camera records, the more difficult and time consuming it will be to create a compelling fake

sequence of preview frames

- manipulation needs to be propagated through all recorded frames

motion sensor data

- (forged) preview needs to be consistent with the recorded camera motion

- general concept, not tied to specific tools presented in this work

Summary

- the more information that a camera records, the more difficult and time consuming it will be to create a compelling fake

sequence of preview frames

- manipulation needs to be propagated through all recorded frames

motion sensor data

- (forged) preview needs to be consistent with the recorded camera motion

- general concept, not tied to specific tools presented in this work
- cameras and mobile devices add more and more sensors, e. g., ambient light and proximity sensors

Summary

- the more information that a camera records, the more difficult and time consuming it will be to create a compelling fake

sequence of preview frames

- manipulation needs to be propagated through all recorded frames

motion sensor data

- (forged) preview needs to be consistent with the recorded camera motion

- general concept, not tied to specific tools presented in this work
- cameras and mobile devices add more and more sensors, e. g., ambient light and proximity sensors

- recording additional data increases image file size (5 seconds preview data at 10 fps $\approx$ 750K)

Summary

- the more information that a camera records, the more difficult and time consuming it will be to create a compelling fake

sequence of preview frames

- ▷ manipulation needs to be propagated through all recorded frames

motion sensor data

- ▷ (forged) preview needs to be consistent with the recorded camera motion

- general concept, not tied to specific tools presented in this work
- cameras and mobile devices add more and more sensors, e. g., ambient light and proximity sensors

- recording additional data increases image file size (5 seconds preview data at 10 fps ≈ 750K)

- approach has limitations for static scenes captured with a stationary camera (on a tripod), or highly dynamic scenes

Summary

- the more information that a camera records, the more difficult and time consuming it will be to create a compelling fake

sequence of preview frames

- ▷ manipulation needs to be propagated through all recorded frames

motion sensor data

- ▷ (forged) preview needs to be consistent with the recorded camera motion

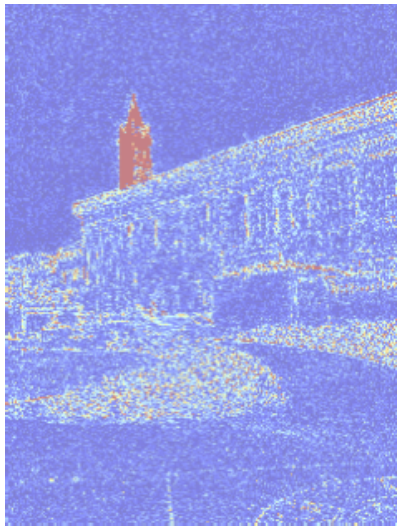
- general concept, not tied to specific tools presented in this work
- cameras and mobile devices add more and more sensors, e. g., ambient light and proximity sensors
- recording additional data increases image file size (5 seconds preview data at 10 fps ≈ 750K)
- approach has limitations for static scenes captured with a stationary camera (on a tripod), or highly dynamic scenes
- unlikely to be vulnerable to a single counter-measure

Impeding Forgers at Photo Inception

Matthias Kirchner [a]
Peter Winkler [b]
Hany Farid [b]

[a] ICSI Berkeley | [b] Dartmouth College





Questions?

What can we (hope to) know?

from specialized hardware ?

- stereo camera, light field camera
- ▷ 3-D scene structure from a multitude of images





What can we (hope to) know?

from specialized hardware?

- stereo camera, light field camera
- ▷ 3-D scene structure from a multitude of images





from external sources?

- Google Street View, Flickr, . . .
- ▷ 3-D scene reconstruction

from metadata (EXIF)?

- ▷ basic auxiliary information, GPS data

What can we (hope to) know?

from specialized hardware?

- stereo camera, light field camera
- ▷ 3-D scene structure from a multitude of images





from external sources?

- Google Street View, Flickr, . . .
- ▷ 3-D scene reconstruction

from metadata (EXIF)?

- ▷ basic auxiliary information, GPS data

from standard consumer devices?







What can we (hope to) know?

from specialized hardware?

- stereo camera, light field camera
  - 3-D scene structure from a multitude of images





from external sources?

- Google Street View, Flickr, . . .
  - 3-D scene reconstruction

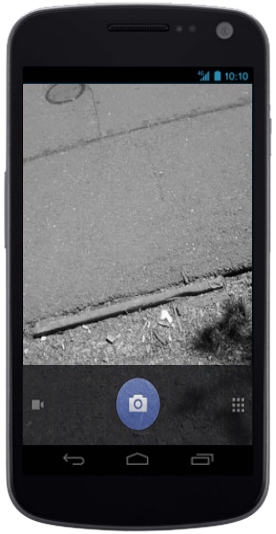
from metadata (EXIF)?

- basic auxiliary information, GPS data

from standard consumer devices?

- most devices provide a digital preview
  - if recorded with the image, manipulations must propagate

- motion sensors
  - 3-D structure of a forged preview must be consistent with recorded camera motion



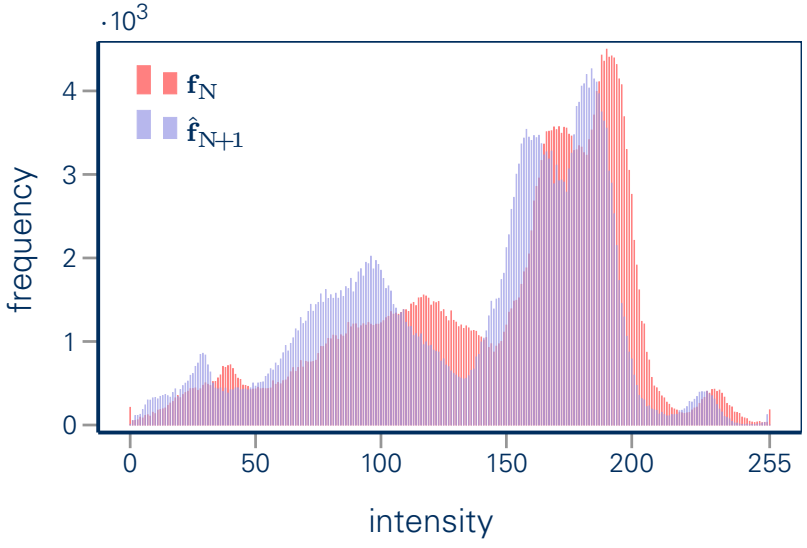


Photometric alignment

- sequential frames $(\mathbf{f}_{t-1}, \hat{\mathbf{f}}_t)$ may vary photometrically
  - auto-exposure, gamma correction, saturation correction

$\mathbf{f}_N$







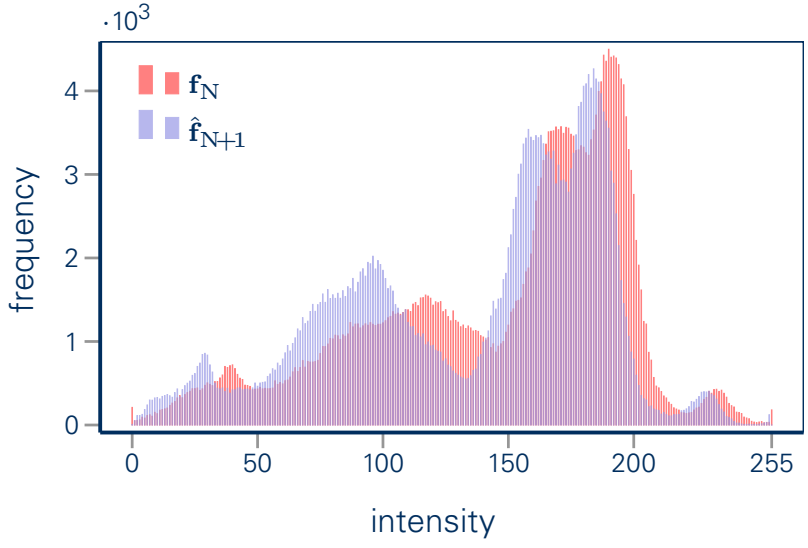
$\hat{\mathbf{f}}_{N+1}$

Photometric alignment

- sequential frames $(\mathbf{f}_{t-1}, \hat{\mathbf{f}}_t)$ may vary photometrically
  - ▷ auto-exposure, gamma correction, saturation correction

$\mathbf{f}_N$



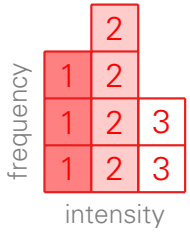


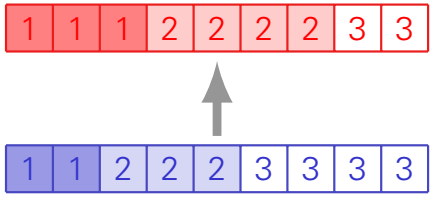


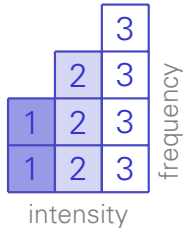
$\hat{\mathbf{f}}_{N+1}$

- histogram matching to yield aligned frames $(\mathbf{f}_{t-1}, \tilde{\mathbf{f}}_t)$ with identical histograms

target histogram







original histogram

Photometric alignment

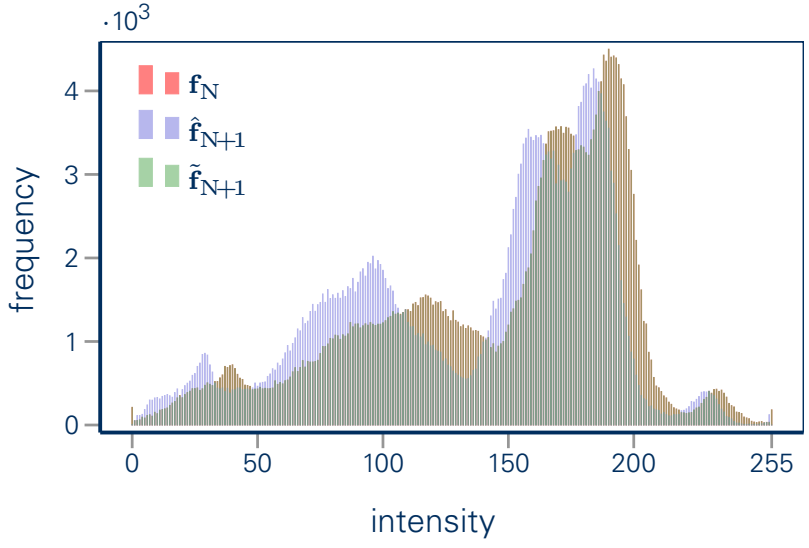
- sequential frames $(\mathbf{f}_{t-1}, \hat{\mathbf{f}}_t)$ may vary photometrically
  - ▷ auto-exposure, gamma correction, saturation correction



$\tilde{\mathbf{f}}_{N+1}$

$\mathbf{f}_N$







$\hat{\mathbf{f}}_{N+1}$

- histogram matching to yield aligned frames $(\mathbf{f}_{t-1}, \tilde{\mathbf{f}}_t)$ with identical histograms

