Restoration of Out-of-Focus Lecture Videos by Automatic Slide Matching

Presented by
Yekaterina Kharitonova

October 1, 2010

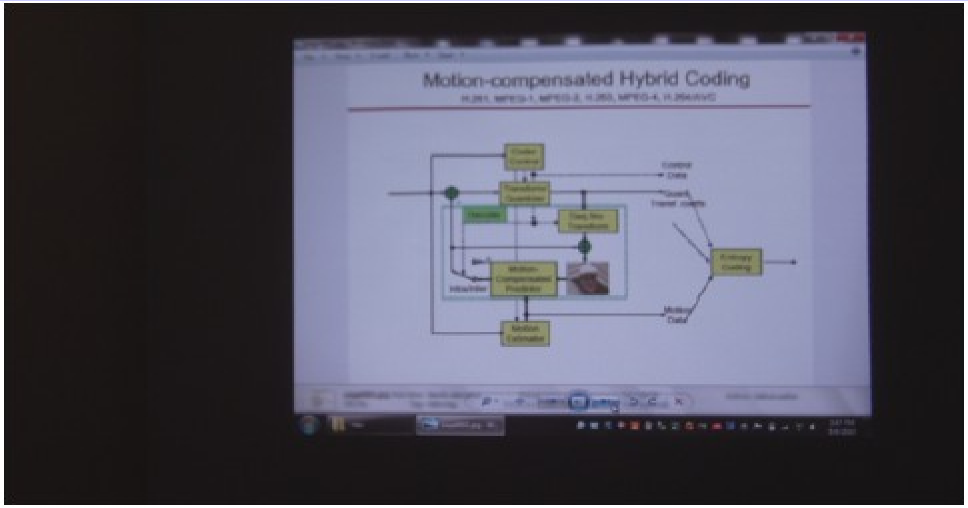
Motivation

- Lecture video capturing
  - Classes
  - Training materials

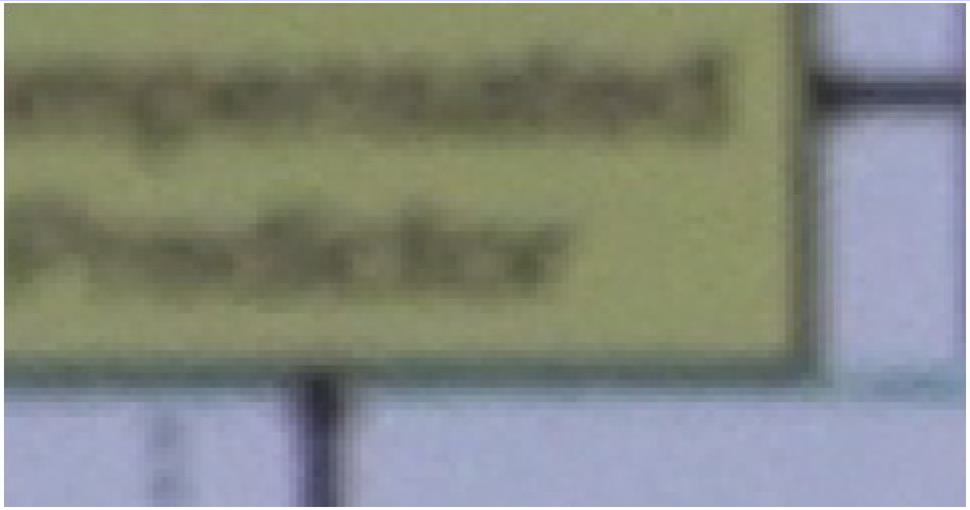
  - Low cost automated systems
  - Non-professional recording

- Challenges
  - Inadequate illumination
  - camera/projector defocusing

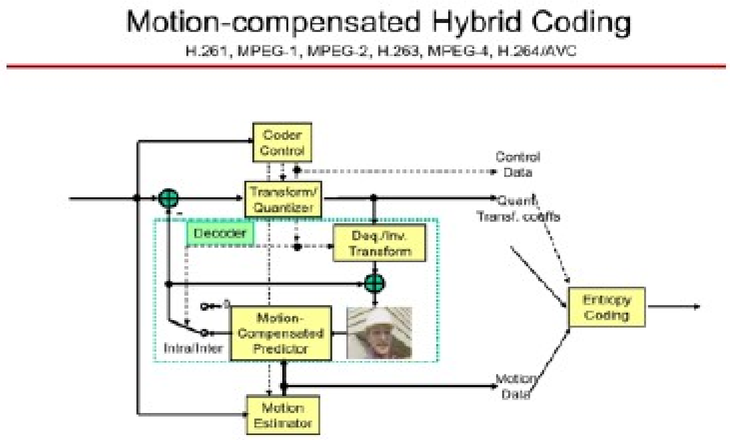
Lecture Video Restoration



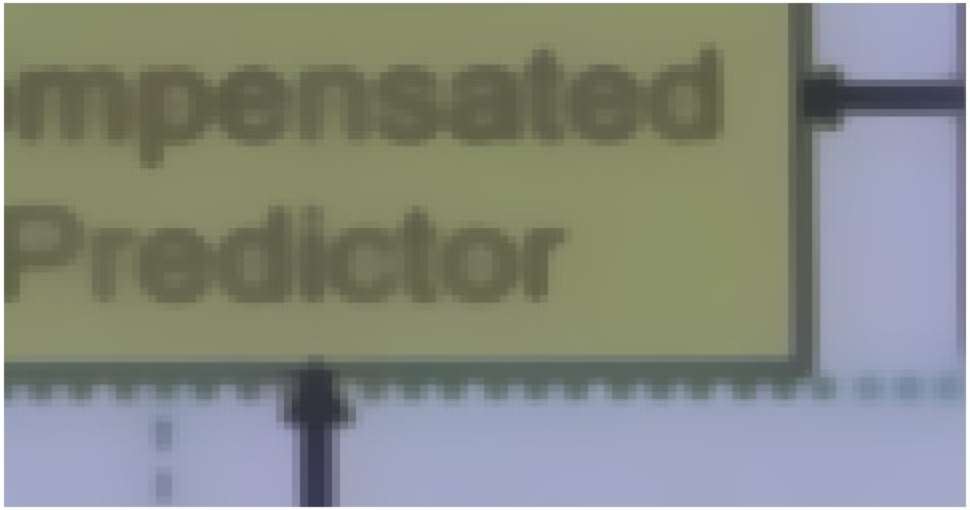
(a) Defocused video frame



(b) Close-up in defocused



(c) Electronic slide



(d) Our restoration result

* All images for this presentation are taken from the original paper “Restoration of Out-of-Focus Lecture Videos by Automatic Slide Matching” by Cheung, Ngai-man, Chen, David et al.

Problem Formulation

- Clean video frame ***a(x, y)***
- Spatially invariant *defocus kernel* ***h(x,y)***
- White zero-mean noise ***n(x,y)***
- Defocused frame ***f(x,y)***

$$f(x,y) = a(x,y) \bullet h(x,y) + n(x,y)$$

- Defocus distortion *h(x,y)* can be well modeled by symmetric circular 2-D Gaussian:

$$h(x,y;\sigma_h) = \frac{1}{2\pi\sigma_h^2} e^{-(x^2+y^2)/2\sigma_h^2}$$

Restoration Algorithm

I. **Automatic slide matching**: match the defocused frame against the slide deck by comparing the local features

II. **Wiener deconvolution**: use the matched slide to estimate the defocus kernel and noise variance

III. **Edge preserving filtering**: apply a bilateral filter to remove the ringing artifacts around sharp edges.

I. Automatic slide matching

Match the defocused frame against the slide deck by comparing the local features

- Get the difference-of-Gaussian filtered images
- Compute 128-dimensional SIFT descriptors
- Establish correspondence between the video frame and the slide image → use RANSAC to establish the projective transformation
- Matching result is the slide that has the max number of correspondences consistent with the estimated geometric transformation

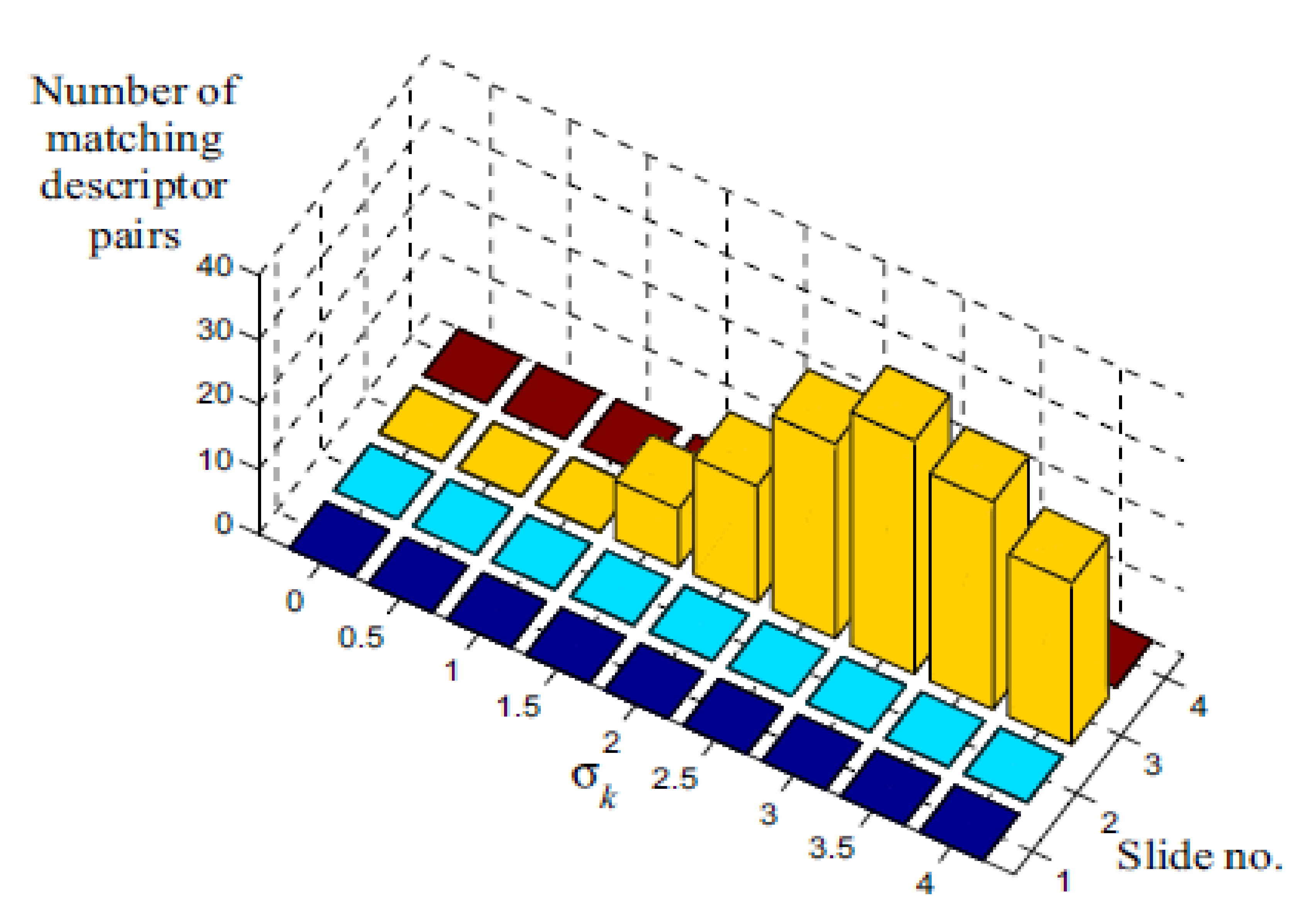
I. Automatic slide matching - Challenges

- Defocus reduces interest point repeatability
  - Reduces geometrical stability of the detected interest points between different images
  - Points detected in the slide are not detected at the corresponding points on the frame
- Blur alters the local gradients and distorts the descriptors

I. Automatic slide matching using defocused slide decks

- Convolve the original slide deck with defocus kernels $h(x, y; \sigma_k)$, where σ_k is the defocus scale (1/2, 1, 3/2, ..., 4)
- Perform pairwise comparison between the blurry video frame and the stack of defocused slide decks
- Declare the match to be the one with max number of matching descriptor pairs

Automatic slide matching with multiple defocused slide decks



II. Wiener deconvolution

Use the matched slide to restore the out-of-focus video frame by estimating the defocus kernel and noise variance

- The spectrum of the sharpened image, i(x, y):

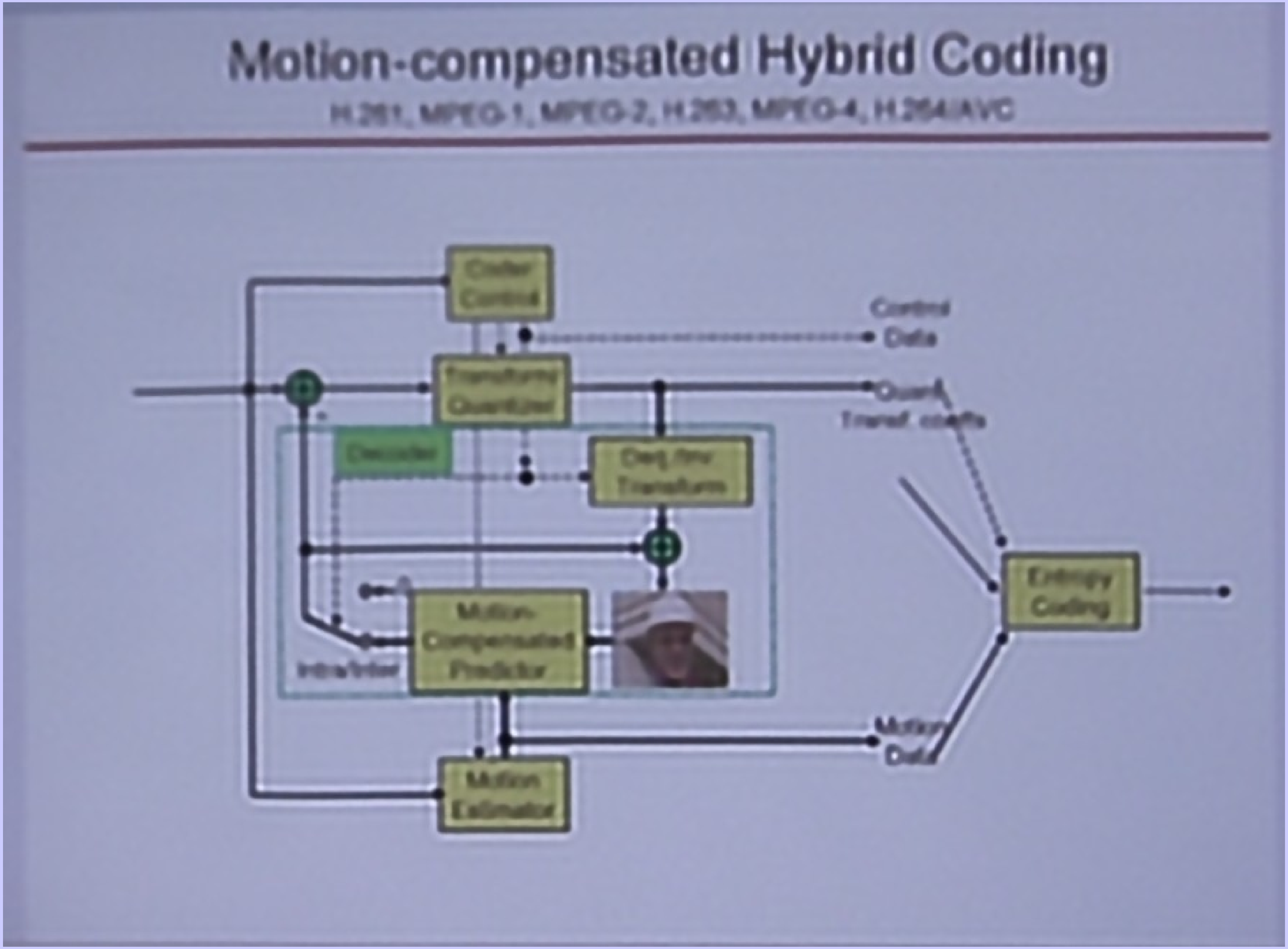
$$I(u,v) = \frac{\Phi_{aa}(u,v)H^*(u,v)F(u,v)}{\Phi_{aa}(u,v)|H(u,v)|^2 + \Phi_{nn}(u,v)},$$

- *H(.)*, *F(.)* are the Fourier transforms of *h(.)*, *f(.)*
- H*(.) is the complex conjugate of H(.)
- Φ_{aa}, Φ_{nn} are the power spectral density of *a(.)* and *n(.)*

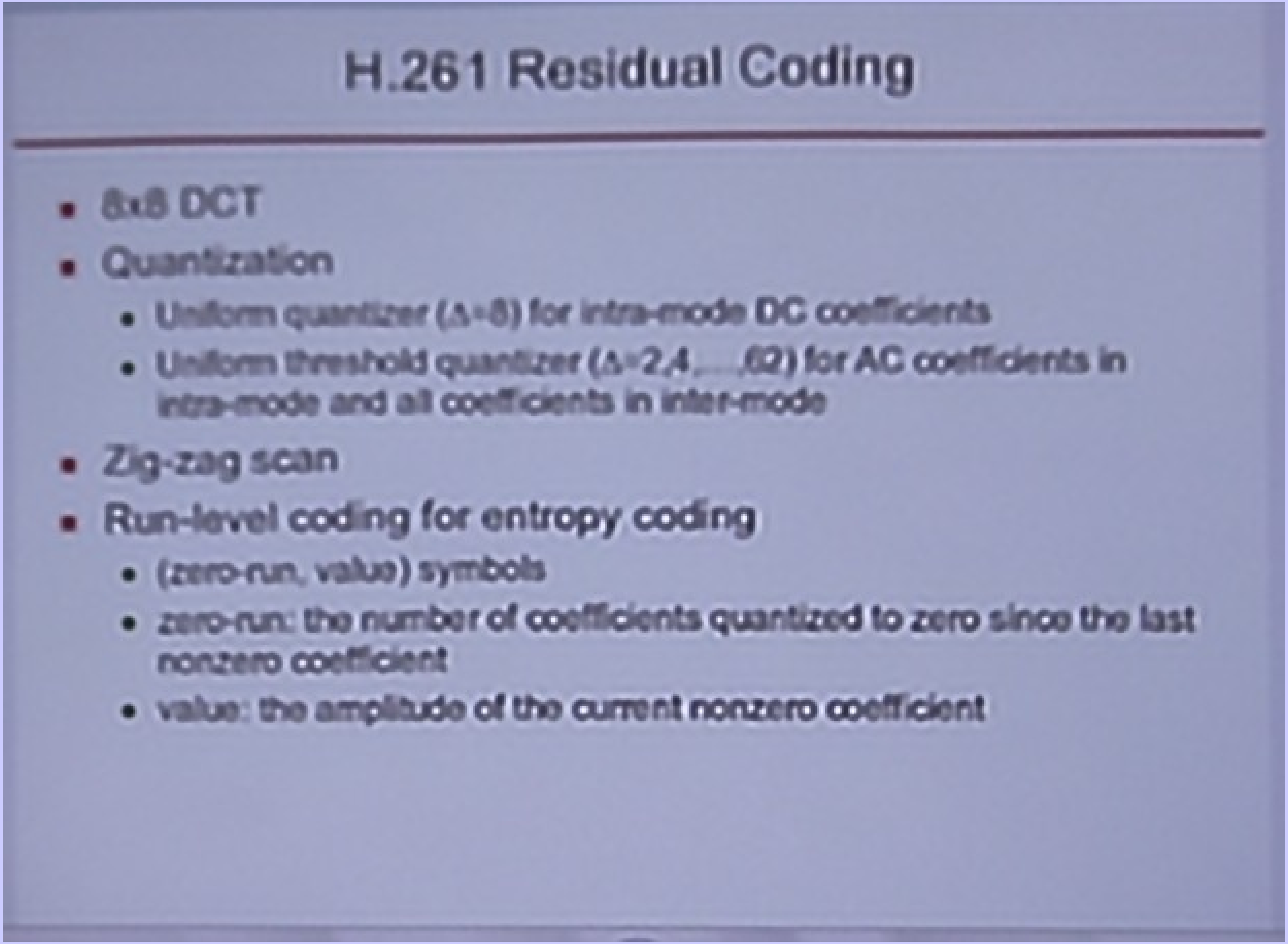
II. Wiener deconvolution

- The spread σ_h can be approximated by the scale σ_k of the defocus kernel that leads to the maximum matches
- To approximate the variance of white noise σ_n
  - Use the computed geometric transformation to backproject the slide into the video frame
  - Partition the projected slide $l(x,y)$ into non-overlapping m x m regions
  - Detect regions of zero variance from $l(x,y)$
  - Compute the variance of the corresponding co-located regions in the defocused frame

Blind Deconvolution



Blind Deconvolution



III. Edge preserving filtering

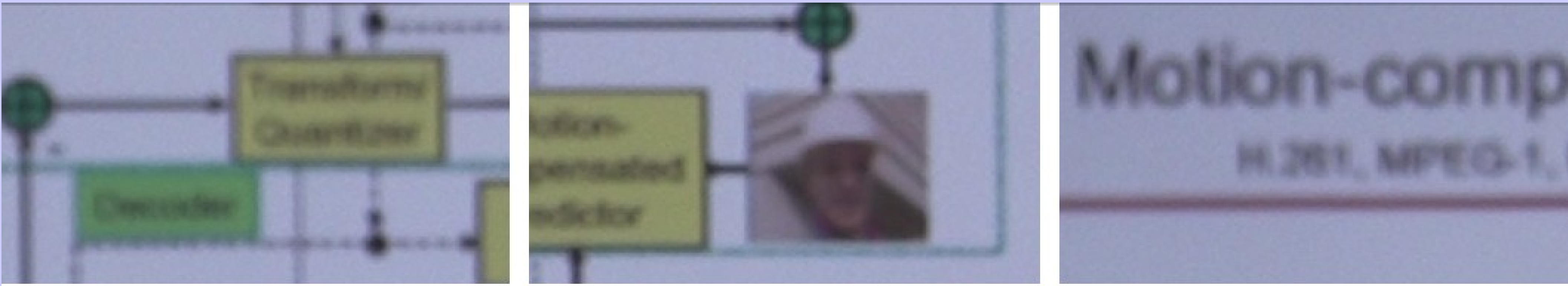
Apply a bilateral filter to remove the ringing artifacts around sharp edges.

- Bilateral filtering of the deconvolution output i(.) is given by:

$$BF[i(.)]_{\mathbf{p}} = \frac{1}{w_{\mathbf{p}}} \sum_{\mathbf{q}} G(\|\mathbf{p}-\mathbf{q}\|;\sigma_s) G(|l(\mathbf{p})-l(\mathbf{q})|;\sigma_r) i(\mathbf{q}).$$

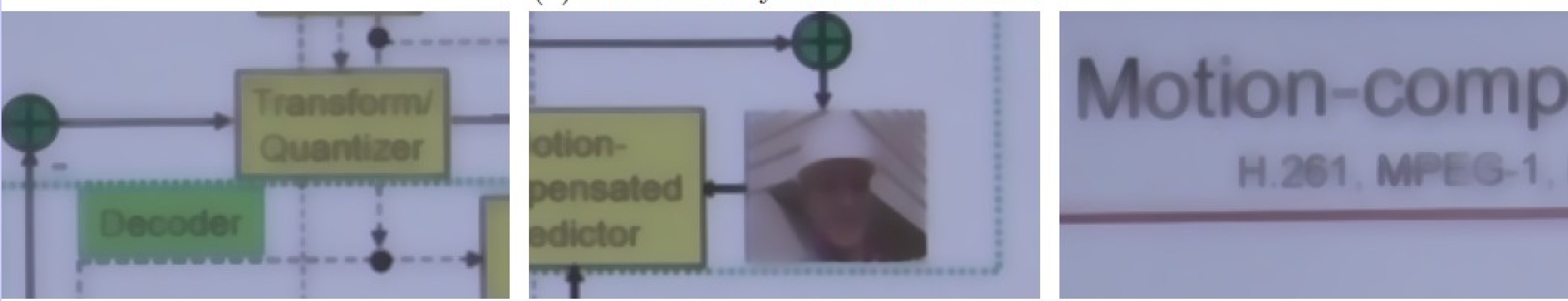
- The intensity value at each pixel in an image is replaced by a weighted average of intensity values from nearby pixels.

Experimental Results



(a) Defocused video frames

(b) Restoration by blind deconvolution



(c) Slide-assisted restoration

Thank you!