G. Cheung

National Institute of Informatics

14th February, 2011

Sparse Representation of Depth Maps for Efficient Transform Coding

.

Outline

- Introduction of NII
- Introduction of my research
- Background to Depth Map Encoding
- 1-slide Summary of Contributions
- Related Work
- Hard Thresholding: Don't Care Region (DCR)
- Soft Thresholding: Penalty Functions
- Conclusion

Outline

- Introduction of NII
- Introduction of my research
- Background to Depth Map Encoding
- 1-slide Summary of Contributions
- Related Work
- Hard Thresholding: Don’t Care Region (DCR)
- Soft Thresholding: Penalty Functions
- Conclusion

Introduction of NII

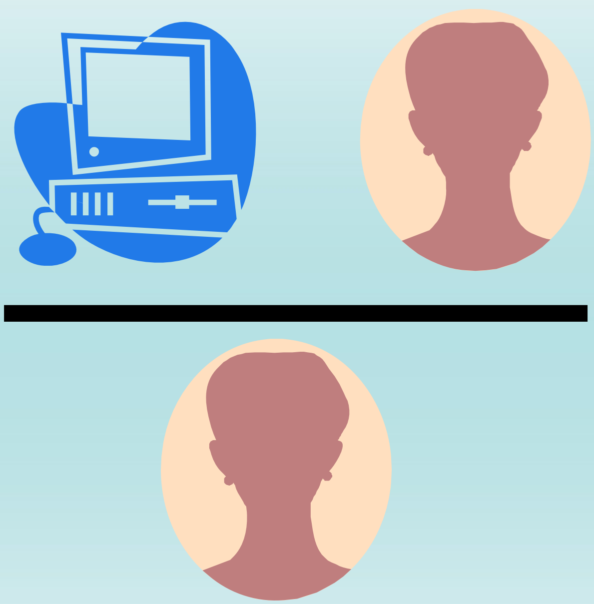
- National Institute of Informatics
- Chiyoda-ku, Tokyo, Japan.
- Fairly new government-funded research lab.
- Offers graduate courses & degrees through The Graduate University for Advanced Studies.
- 60+ faculty in “informatics”: quantum computing, discrete algorithms, machine learning, computer networks, computer vision, image & video processing.
- Foreigner-friendly, actively seeking int’l collaborations.

Outline

- Introduction of NII
- Introduction of my research
- Background to Depth Map Encoding
- 1-slide Summary of Contributions
- Related Work
- Hard Thresholding: Don’t Care Region (DCR)
- Soft Thresholding: Penalty Functions
- Conclusion

Turing Test

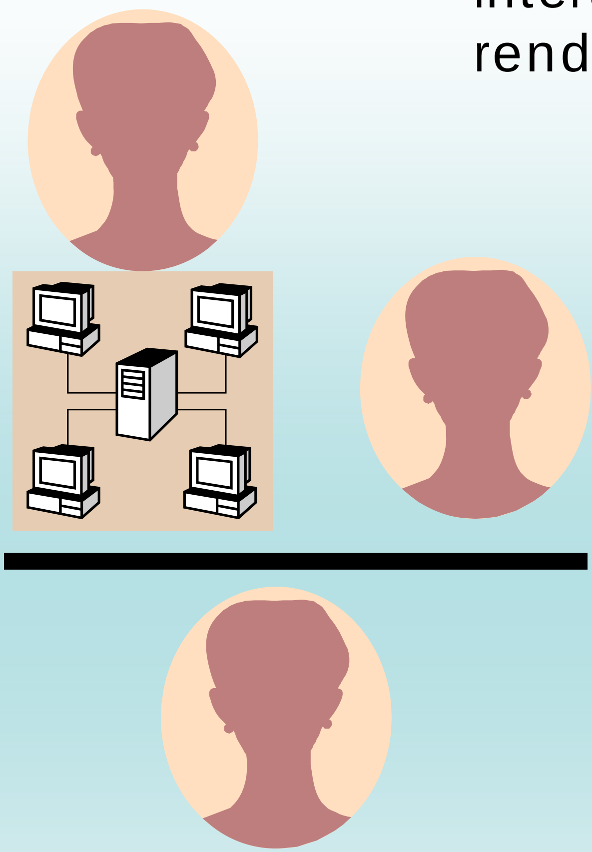
- Alan Turing introduced test in 1950.
- **Q**: can a person engage in natural language conversation, and not be able to tell if participant is computer or human?



A. Turing, “Computing Machinery and Intelligence”, *Mind*, (236): 433–460, Oct, 1950.

Immersive Experience Test

- **Q**: can a person engage in natural inter-personal interaction, and not be able to tell if participant is rendered images or actual human?



Large display w/ HQ life-size images

Gaze-corrected view

Motion Parallax: Fast view-switching via head tracking

Multiview video coding & View Synthesis

Loss/delay tolerant multiview transmission

Natural visual media interaction

Outline

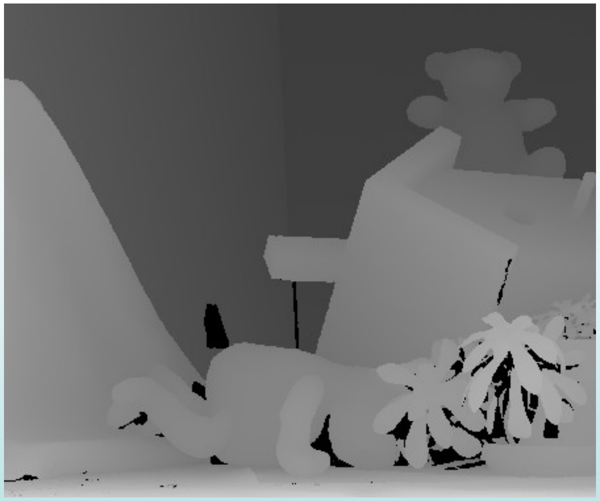
- Introduction of NII
- Introduction of my research
- Background to Depth Map Encoding
- 1-slide Summary of Contributions
- Related Work
- Hard Thresholding: Don’t Care Region (DCR)
- Soft Thresholding: Penalty Functions
- Conclusion

Background to Depth Map Encoding

- Multiview Imaging:
  - Closely spaced cameras taking pictures simultaneously.
  - Besides captured **texture maps**, **depth maps** can also be captured / estimated.
  - Texture / depth maps enable synthesis of intermediate views using **Depth-Image-Based Rendering** (DIBR).
  - Also called "Image / video + depth" format.

- Depth Map Compression Problem:
  - How to efficiently encode depth maps in a rate-distortion optimal way?

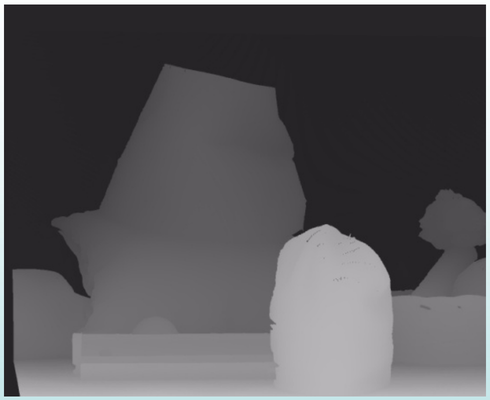
1-slide Summary of Contributions



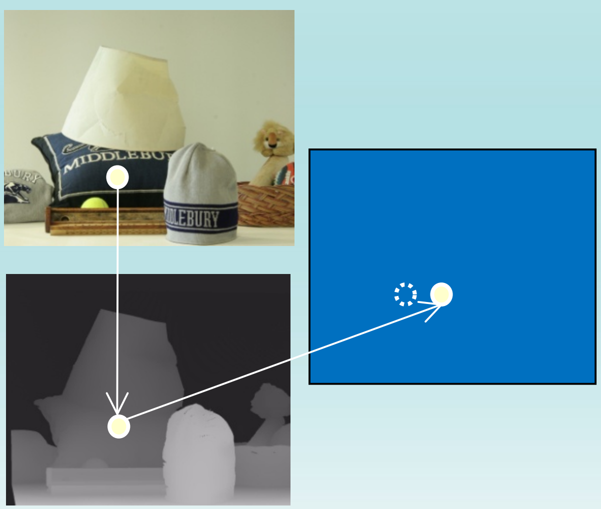
- **Key Observation**:
  - Depth map is:
    - NOT for direct observation.
    - For interpolation of intermediate views via DIBR.
  - ***Can manipulate depth values WITHOUT directly causing visual distortion.***

- **Key Idea**: *sparse transform coding*
  1. Define **per-pixel sensitivity** for depth map according to its effect on DIBR.
  2. Find sparse rep. in transform domain for compression gain, given per-pixel sensitivity.

Related Work



- *Depth Map Specific Compression*
  - Depth characteristics: smooth surface & sharp edges.
  - Edge encoding + adaptive wavelet [Maitre TIP'08].
  - Diff: We manipulate depth value directly for compression gain.



- *Depth Map Distortion Analysis*
  - Depth err → position err → copy wrong texture pixel.
  - New metric for block-by-block mode selection [Kim ICIP'09].
  - Diff: We manipulate depth values given defined error sensitivity for sparsity in trans. coding.

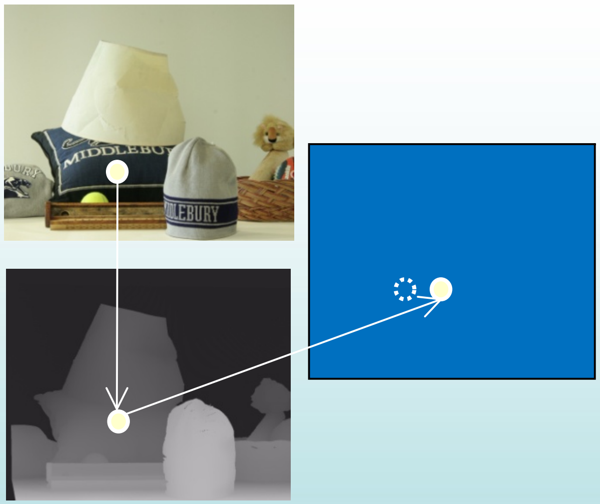
Related Work

- *Signal manipulation in decoded JPEG*
  - Indep. DCT block transform → high freq. boundaries.
  - Signal in quan bins w/o HF via POCS [Rosenholtz CSVT'92].
  - Diff: We manipulate depth values in pixel domain, to maximize sparsity in trans. coding.
- *Signal manipulation in LBT coding*
  - distortion vs. l_1-norm of trans. coeff. [Winken ICIP'10].
  - Diff: diff. DCRs for diff. depth pixels due to DIBR.
  - Diff: sparsity in trans. domain → l_0 minimization.

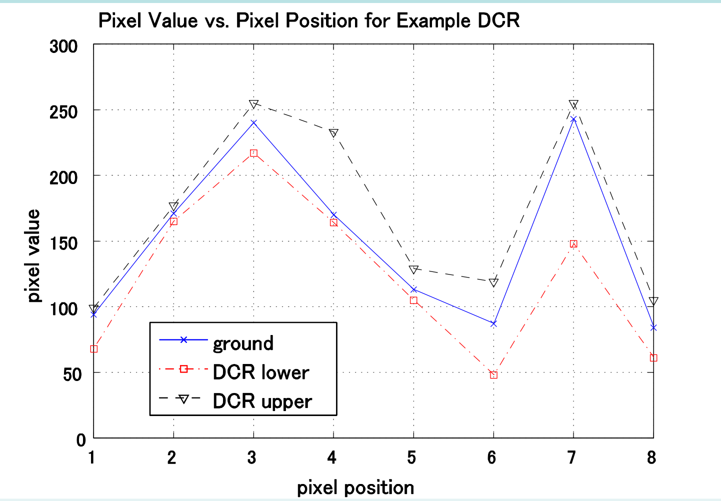
Outline

- Introduction of NII
- Introduction of my research
- Background to Depth Map Encoding
- 1-slide Summary of Contributions
- Related Work
- Hard Thresholding: Don't Care Region (DCR)
- Soft Thresholding: Penalty Functions
- Conclusion

Don't Care Region (DCR)



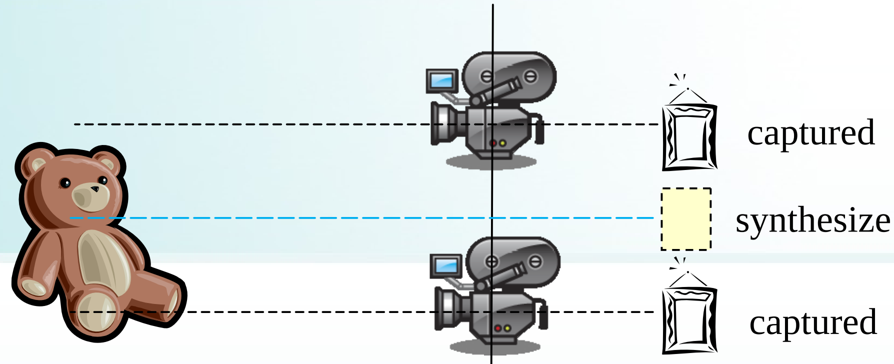
- What is DCR?
  - DCR of pixel (i,j) = range of depth values, s.t. err of synth pixel value ≤ pre-defined threshold.
  - **Intuition**: DCRs larger in smooth textural regions.
  - **Note**: unique to depth maps, not done in literature!



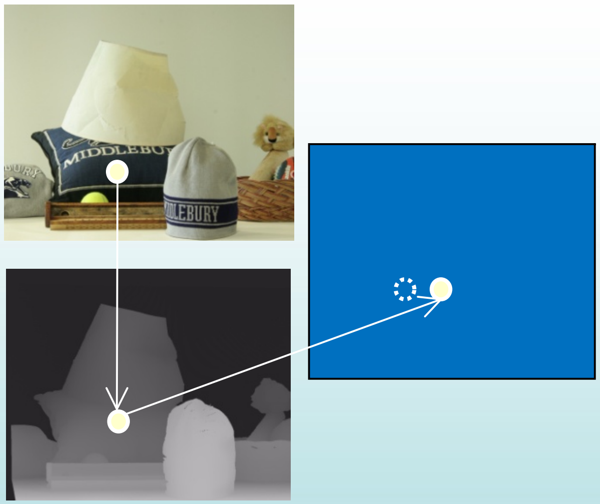
- Key Questions:

1. How to formally define DCR?
2. Given DCR, how to find sparse rep in compressed domain?

Don’t Care Region



- System Setup:
  - Two horizontally shifted cameras.
  - Interpolate middle view w/ middle depth map only.
  - Encode middle depth map only.

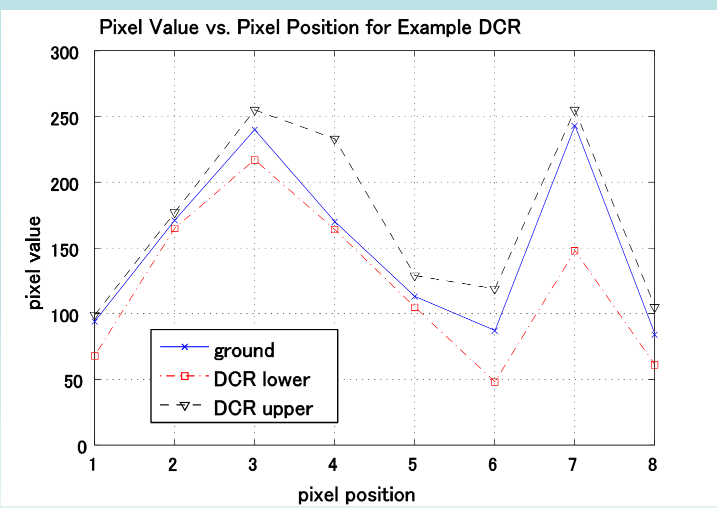


- Derive ground truth depth map:
  - Synthesize middle view with left & right texture maps:

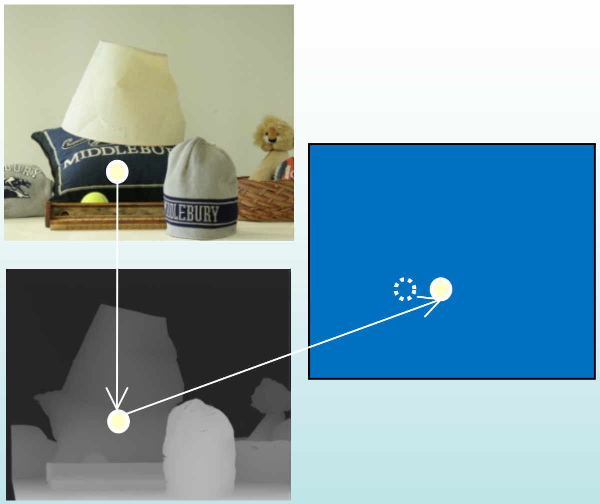
$$I'_{mid}(i,j;d) = \frac{1}{2} I_{left}(i+d,j) + \frac{1}{2} I_{right}(i-d,j)$$

  - Ground truth is depth value w/ smallest err:

$$d_{\min}(i,j) = \arg\min_{d} \underbrace{\left| I'_{mid}(i,j;d) - I_{mid}(i,j) \right|}_{e(i,j;d)}$$



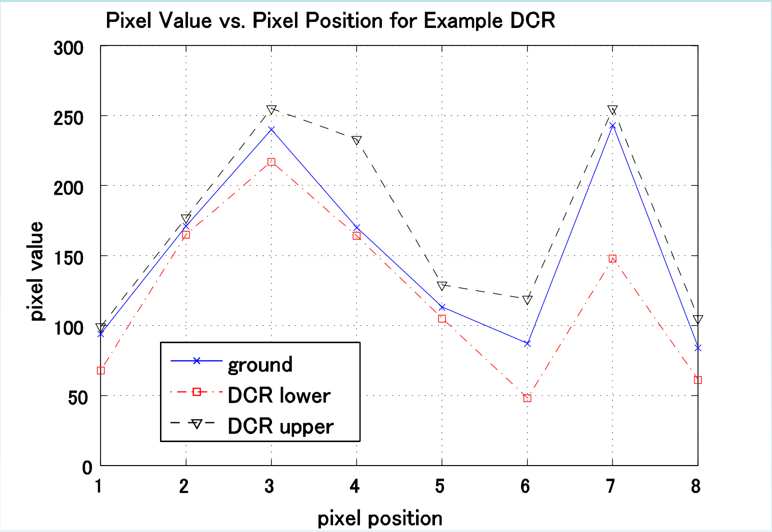
Don't Care Region



- Derive DCR:
  - Define threshold *T*.
  - Find *f(i,j)* and *g(i,j)* around ground truth $d_{min}(i,j)$:

$$f(i,j) = \min\{d\} \xrightarrow{\text{s.t.}} \underbrace{\left| I'_{\min}(i,j;d) - I_{\min}(i,j) \right|}_{e(i,j;d)} \leq T$$

$$g(i,j) = \max\{d\} \xrightarrow{\text{s.t.}} \underbrace{\left| I'_{\min}(i,j;d) - I_{\min}(i,j) \right|}_{e(i,j;d)} \leq T$$



- Large threshold *T*,
  - large search space for sparse rep. in transform domain.
  - large err in synthesized distortion.

Problem Formulation

- Given DCR *R*,
  - find *s* in *R* with sparse rep. *a* in transform domain:

signal in transform domain

orthogonal transform

within defined DCR:

$f_{i,j} \le s_{i,j} \le g_{i,j}$

$$\min_{s \in R} \left\| \vec{a} \right\|_{l_0} \quad \text{s.t.} \quad \vec{a} = \Phi \vec{s}$$

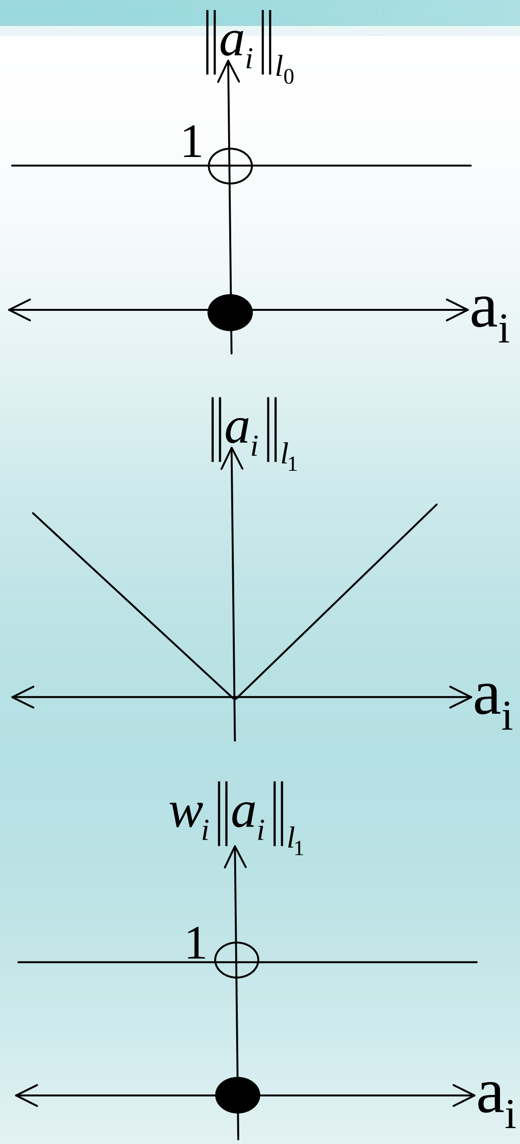
signal in pixel domain

- l_0 norm is # of non-zero coeff's.

$$\left\| \vec{a} \right\|_{l_0} = \left| \left\{ i : a_i \neq 0 \right\} \right|$$

- Combinatorial, difficult to solve.

Surrogate Objective



linear objective function

- Given l_0 is hard, solve l_1 (surrogate) instead.

$$\min_{\vec{s}\in R} \|\vec{a}\|_{l_1} = \sum_i |a_i| \quad \text{s.t.} \quad \vec{a} = \Phi\vec{s}$$

$f_{i,j} \le s_{i,j} \le g_{i,j}$

linear constraints

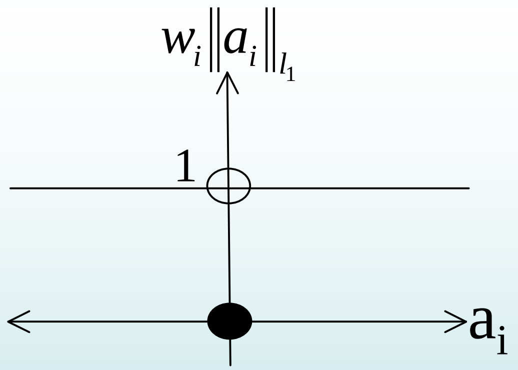
- Efficiently solved via *linear programming*.
- l_1 is quite different from l_0 , so weighted l_1?

$$\|\vec{a}\|_{l_1^w} = \sum_i w_i |a_i|$$

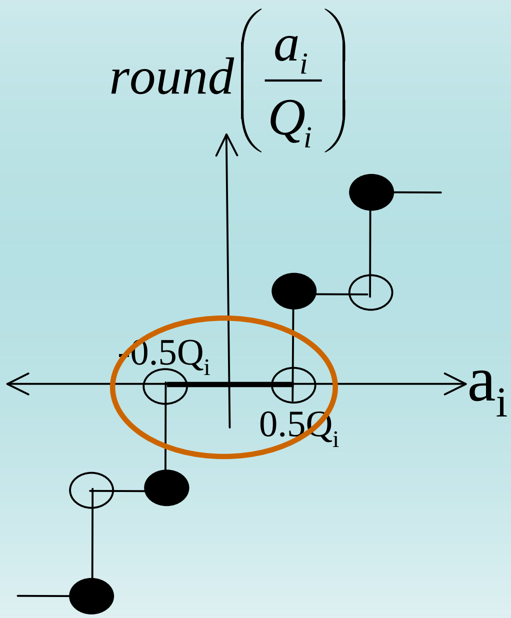
w_i = $1/|a_i|$ if $a_i \neq 0$,
= 0 o.w.

- Problem: don't know weights $1/|a_i|$'s a priori.

Surrogate Objective



- Sol'n: iterative algorithm*
  1. Init weights $w_i = 1$.
  2. Solve l_1 minimization for sol'n a_i's.
  3. Set weights $w_i = 1/|a_i|$.
  4. Repeat step 2 and 3 till convergence.

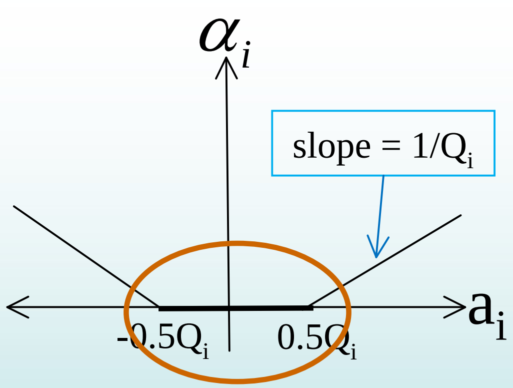


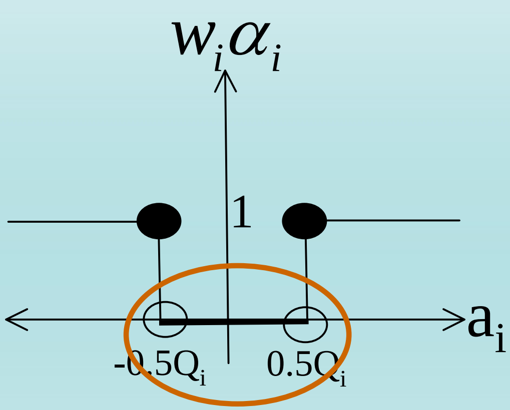
- Actually, want sparse ***quantized*** coeff.
  - Quant coeff = $round\left(\frac{a_i}{Q_i}\right)$
  - Non-zero quant coeff only if $\left|\frac{a_i}{Q_i}\right| \geq 0.5$

*Candes et al., "Enhancing sparsity by reweighted l_1 minimization," *JFAA*, 12/2008.

Surrogate Objective

Recall:
non-zero quant coeff only if $\left|\frac{a_i}{Q_i}\right| \geq 0.5$





- Define shrinkage coeff:

$$\alpha_i = \max\left\{\left|\frac{a_i}{Q_i}\right| - 0.5,\, 0\right\}$$

- Define new obj. func: $\min \sum_i w_i \alpha_i$

linear objective function

- Write α_i in linear form:

$$\alpha_i \geq \frac{a_i}{Q_i} - 0.5$$

$$\alpha_i \geq 0$$

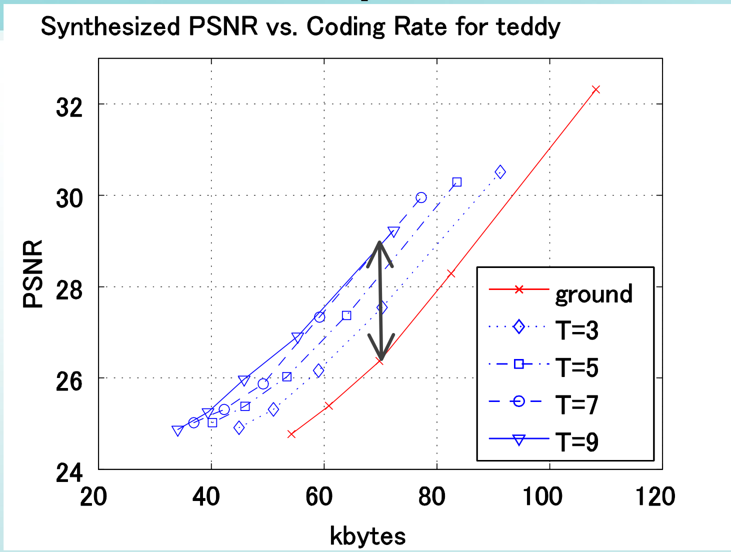
$$\alpha_i \geq -\frac{a_i}{Q_i} - 0.5$$

$$\vec{a} = \Phi \vec{s}$$

$$f_{i,j} \leq s_{i,j} \leq g_{i,j}$$

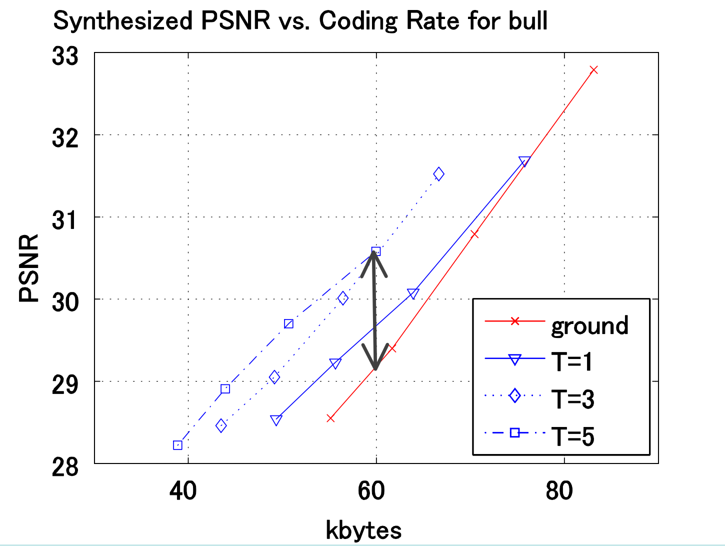
linear constraints

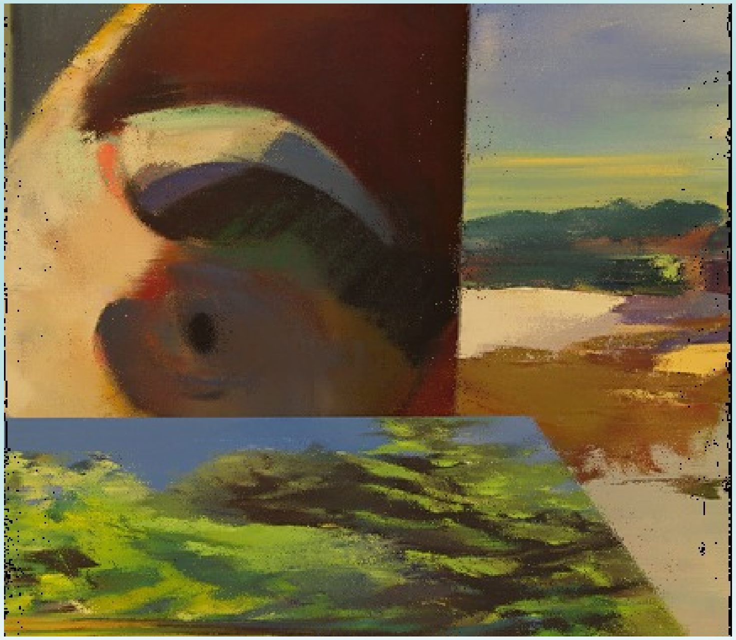
Experimental Results #1



- Independent JPEG grroup's cjpeg version 8a.
- Multiview seq. teddy from Middlebury.
- Optimize 8x8 pixel block at a time.
- Fixed Threshold T, opt all blocks of depth map and vary QP.
- Texture maps not compressed.
- Observations:
  1. As T increases, RD performance improves.
  2. Up to 2.5dB improvement of ground truth.
  3. No annoying visual artifacts due to sparse representation.

Experimental Results #2





- Multiview seq. bull from Middlebury.

- Observations:
  1. As T increases, RD performance improves, but improvement tails off faster.
  2. Up to 1.5dB improvement of ground truth.
  3. No annoying visual artifacts due to sparse representation.

Outline

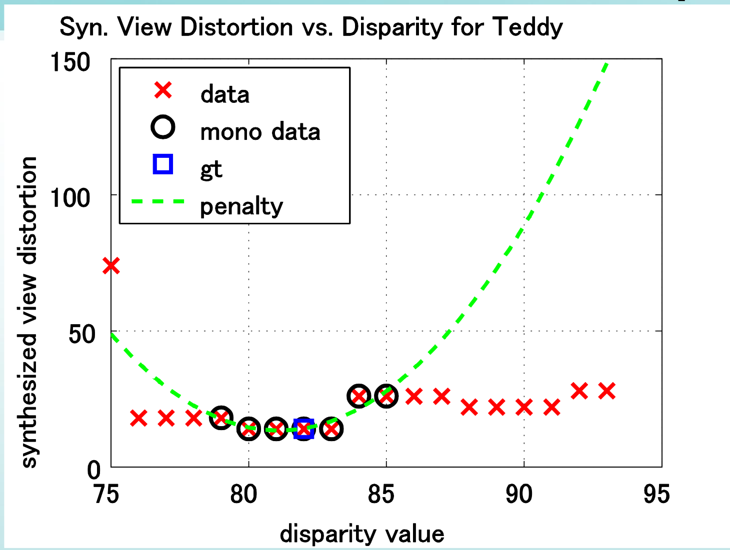
- Introduction of NII
- Introduction of my research
- Background to Depth Map Encoding
- 1-slide Summary of Contributions
- Related Work
- Hard Thresholding: Don't Care Region (DCR)
- Soft Thresholding: Penalty Functions
- Conclusion

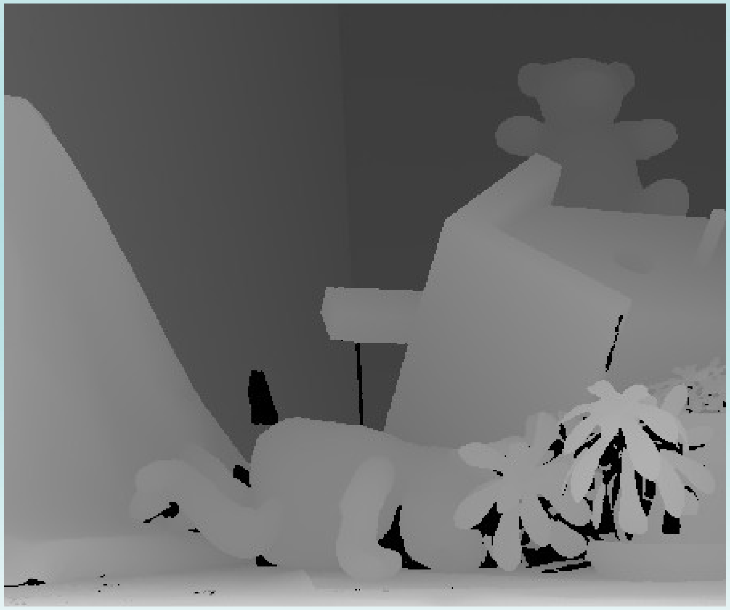
Soft Thresholding

- Problems with Hard Thresholding (DCR):
  1. Optimize for 1 depth map.
  2. Iterative LP still computation expensive.

- Define per-pixel penalty function.
- Promote sparsity using weighted l2-norm.
  - Unconstrained quadratic programs.

Define Per-pixel Penalty Function





- Define quadratic penalty function:

$$g_i(s_i) = \left(\tfrac{1}{2}\right) a_i s_i^2 + b_i s_i + c_i$$

- Synthesized distortion sensitive to depth pixel → sharper parabola.

$$\underbrace{E_l}_{\text{error}}(k;m,n) = \left|\underbrace{I_l}_{\text{left texture map}}\left(m + \underbrace{D_l(m,n) + k}_{\text{shift}}, n\right) - \underbrace{I_r}_{\text{right texture map}}(m,n)\right|$$

Objective Function

$$s = \sum_i \alpha_i \phi_i$$

depth signal

trans. coeff.

basis func.

- Sum of l0-norm + weighted penalties (transform domain):

$$\min_{\alpha} \|\alpha\|_{l_0} + \lambda \sum_i g_i\left(\phi_i^{-1}\alpha\right)$$

- Replace l0-norm with weighted l2-norm:

quadratic
penalty func.

$$\min_{\alpha} \sum_i w_i \alpha_i^2 + \lambda \sum_i g_i\left(\phi_i^{-1}\alpha\right)$$

- Unconstrained quadratic program, solvable via set of linear equations.

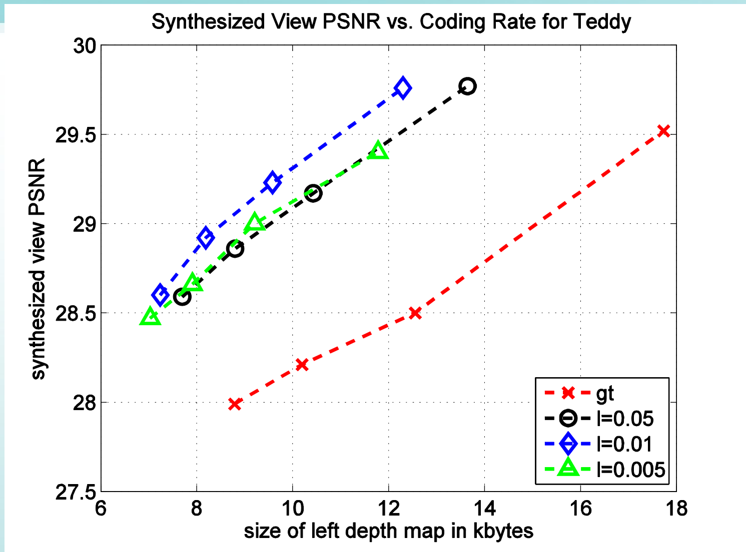
Iterative Quadratic Minimization

1. Init weights $w_i = \left(\left| \alpha_i^t \right|^2 + \varepsilon^2 \right)^{-1}$, where α_i^t is coeff of ground truth depth signal.
2. Find optimal α^o using surrogate objective.
3. Set weight α^o to $\left(\left| \alpha_i^o \right|^2 + \varepsilon^2 \right)^{-1}$ if $\left| \frac{\alpha_i^o}{Q_i} \right| \geq 0.5$ and $\varepsilon^{-\frac{1}{2}}$ o.w.
4. Repeat until convergence.

- Initialize weights using depth signal.
- Discount contribution to weighted l2-norm if quantized to 0.

*Daubechies et al., "Iteratively reweighted least squares minimization for sparse recovery," *Comm. Pure Appl. Math*, 2009.

Experimental Results



- Independent JPEG grroup's cjpeg version 8a.
- Multiview seq. teddy from Middlebury.
- Optimize 8x8 pixel block at a time.
- Texture maps not compressed.
- Observations:
  1. Up to 1.5dB improvement of ground truth.
  2. No annoying visual artifacts due to sparse representation.

Outline

- Introduction of NII
- Introduction of my research
- Background to Depth Map Encoding
- 1-slide Summary of Contributions
- Related Work
- Hard Thresholding: Don’t Care Region (DCR)
- Soft Thresholding: Penalty Functions
- Conclusion

Conclusion & Future Work

- Depth map compression for DIBR.
- Fixed transform, signal manipulation approach:
  1. Define error sensitivity for each depth pixel,
  2. Find most sparse rep. in compressed domain given defined per-pixel error sensitivity.
- Solve weighted l1, l2 surrogate of l0-norm minimization.
- Significant RD improvement in synthesized view.

- Future Work:
  1. Motion-compensated video?

Quantization Effects on DCR

- Quant. in non-zero coeff not accounted for.
- Quant. can force LP-solved sol'n outside DCR.
- Heuristic: 1 more LP to force sol'n inside DCR.