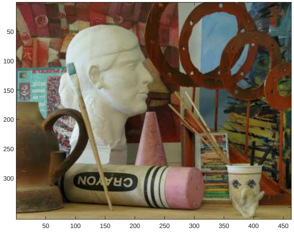
# Modeling and Compression of 3D Image Data

Ioan Tabus

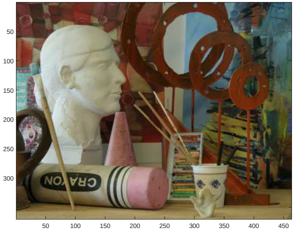
Department of Signal Processing
Tampere University of Technology
Finland

- AREA: Multiview image analysis, processing, and compression
- Essential terms: disparity, depth, warping, left-right redundancy
- Representations of depth information suitable for encoding
- Lossless compression of depth images
- Encoding of L depth image conditional on R depth image
- Lossless compression of color views using warped image for sparse predictive coding
- Applications exploiting the left-right redundancy: Low level image segmentation, Stereo matching, Segmentation of objects

LEFT VIEW $L(i,j)$




RIGHT VIEW $R(i,j)$




$$L(i,j) \approx R(i, j - D^L(i,j)) \qquad R(i,j) \approx L(i, j + D^R(i,j))$$

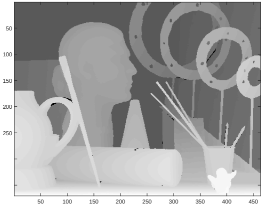
Left image can be seen as a transformed version of Right image
- Warping = shift the argument $j$ by a disparity, which is also a function of coordinates $(i,j)$
- Warping in audio: 1-D transform $y(t) = x(\tau(t))$

# Disparity images: correspondance of points left-right

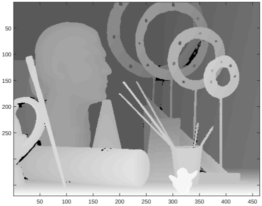
LEFT VIEW $L(i,j)$


RIGHT VIEW $R(i,j)$


LEFT DISPARITY $D^L(i,j)$



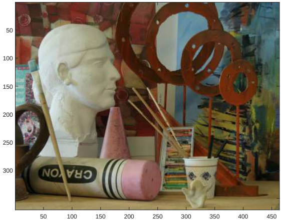

RIGHT DISPARITY $D^R(i,j)$



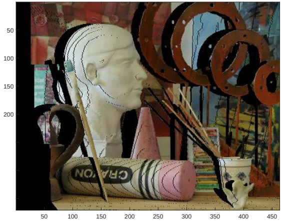

$$D^L(i,j) = D^R(i,j - D^L(i,j)) \quad D^R(i,j) = D^L(i,j + D^R(i,j))$$

- Construct the WARPED IMAGE $L(i, j + D^R(i,j)) \approx R(i,j)$

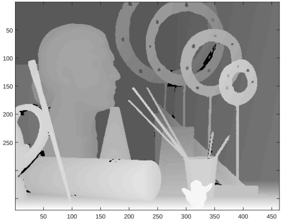
Right Color




Warped image

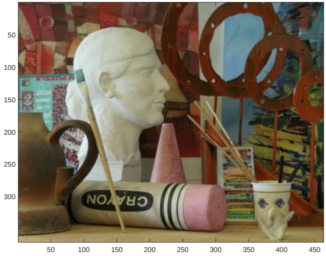



Right Disparity

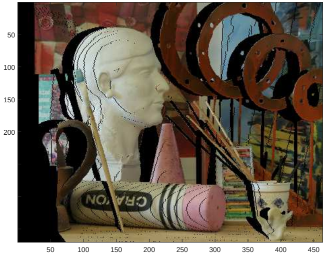


# Compare Left color image to the Right warped color image

LEFT COLOR

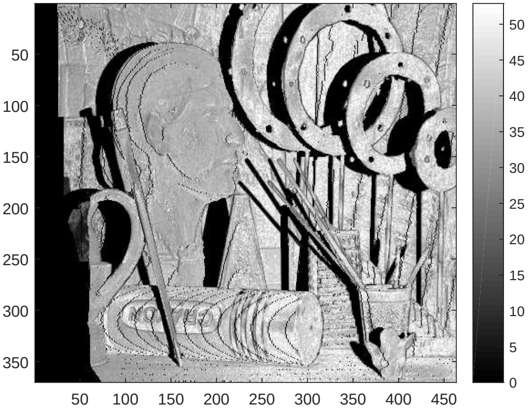



RIGHT WARPED



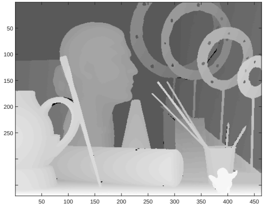

- Warping captures a lot of redundancy
- Warping is not defined everywhere
- Warping does not give a very perfect reconstruction, even where well defined

Reconstruction PSNR(i,j) $= 10\log_{10}(255^2/\varepsilon(i,j)^2)$




- Warping does not give a very perfect reconstruction, even where well defined

- Representations of depth information suitable for encoding
- Lossless compression of depth images
- Encoding of L depth image conditional on R depth image
- Lossless compression of color views using warped image for sparse predictive coding

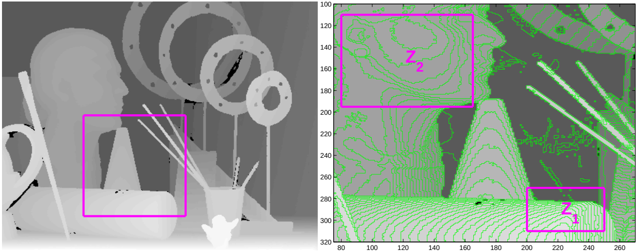



- Disparity $D(i,j)$ and depth $Z(i,j)$ are equivalent

$$D(i,j) = \frac{Bf}{Z(i,j)}$$

$B$ is the baseline (distance between right and left camera centers), $f$ is the focal distance.

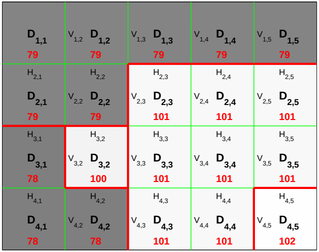
- Depth can be acquired using depth sensors or by stereo matching
- Depth (disparity) images are very smooth, with large constant regions




- right depth image "art" has 4306 constant regions
- REPRESENTATION OF THE IMAGE:
  - Contours of constant regions $\Gamma_1, \ldots, \Gamma_N$
  - Depth value inside each region $d_1, \ldots, d_N$

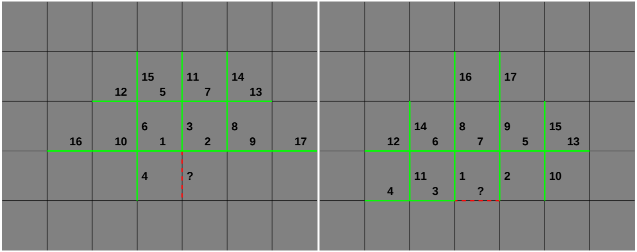
Algorithm CERV:

[1] I. Tabus, I. Schiopu, J. Astola, Context coding of depth map images under the piecewise-constant image model representation. IEEE Trans. Image Processing, 22:11, pp. 41954210, Nov. 2013.

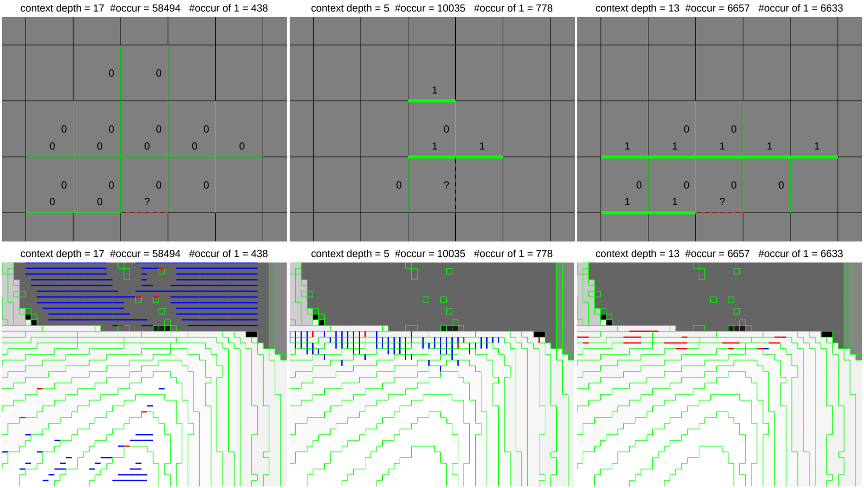


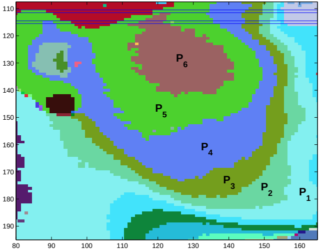

- Two binary images: $\mathcal{H}$ for horizontal crack-edges and $\mathcal{V}$ for vertical crack-edges
- Encode each image using the Context tree algorithm, with bidimensional templates
- The distribution at the contexts is very skewed in most of the contexts
- Design an optimal context tree by dynamic programming pruning
- Transmit the shape of the optimal tree as side information to the decoder

# Examples of templates for the context in $\mathcal{H}$ and $\mathcal{V}$

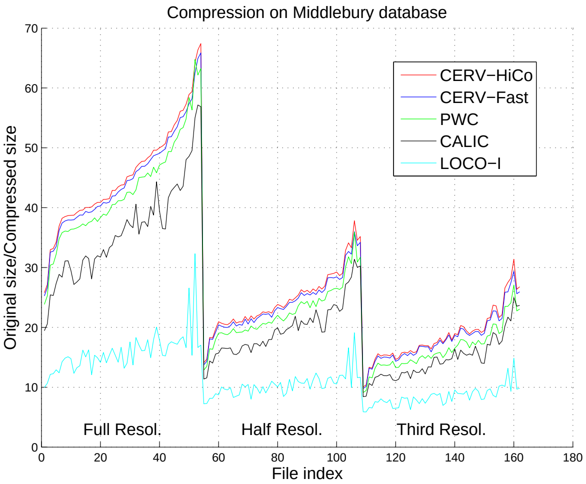


# Examples of pruned contexts in $\mathcal{H}$ and $\mathcal{V}$

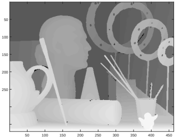





- The patches are forming a graph of neighbor regions
- **Similarity**: neighbor patches are very likely to have similar depth
- **Exclusion**: neighbor patches differ at least by one
- Utilize a list data structure for enforcing exclusions and for collecting statistics of similar values at the neighbors
- Extremely efficient, requires only 10% of the bitrate, the rest is needed for encoding the crack-edges

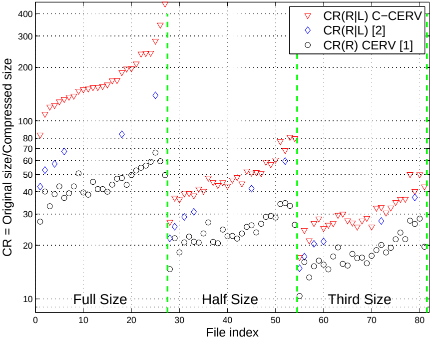



Depth Image $D$

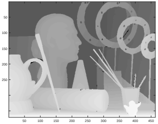


$\mathcal{L}(D)$

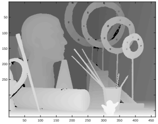
Comparison of the compression ratio (original size over compressed size) for CERV algorithms, PWC, CALIC, and LOCO-I, over the




Encode $D^L$



Given $D^R$



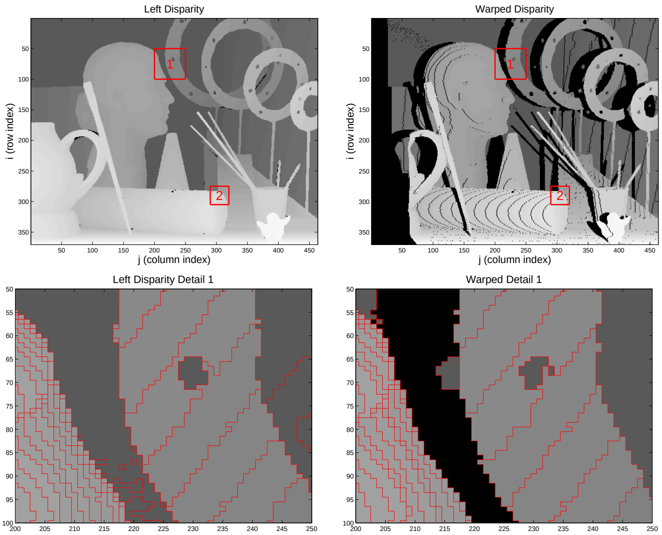
$\mathcal{L}(D^L|D^R)$

.

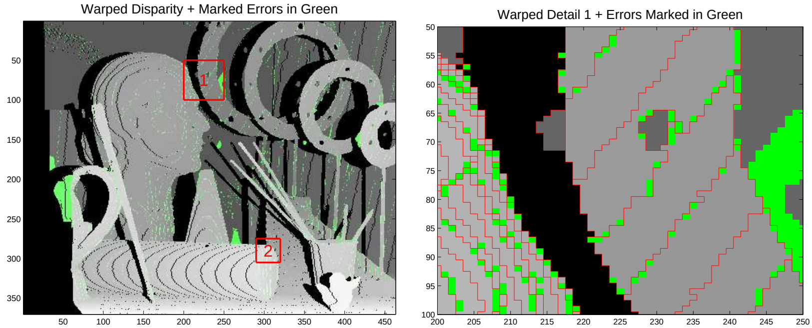
Compression ratios when encoding conditionally using the conditional method for all files "disp1" (left view) and "disp5" (right view) in the Middlebury dataset

[2] I.Tabus, Patch-Based Conditional Context Coding of Stereo Disparity Images. IEEE Signal Processing Letters, 21:10, pp. 12201224, Oct. 2014

# Conditional Context Coding of Stereo Disparity Images: Using Warping

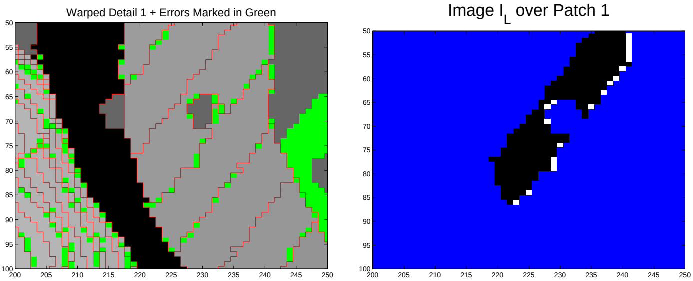


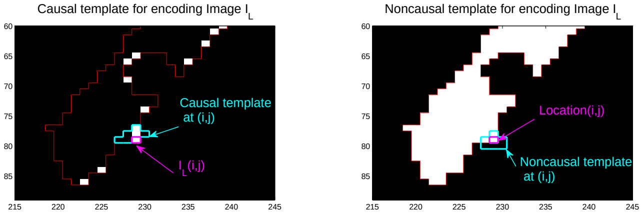

Results of warping: Side-by-side views of the left disparity $D^L$ and the warped disparity $L^w$, obtained when warping $D^R$.

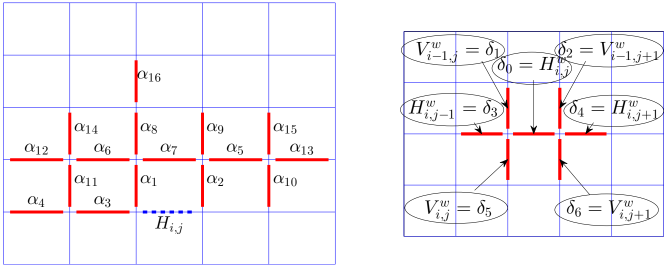



Results of warping: Marking in green the pixels $(i, j)$ where the warped image is incorrect $(L(i,j) \neq L^w(i,j))$ and the warped image is nonzero, $L^w(i,j) > 0$.

The patch contours warped from $D^R$ are known at decoder. The binary image of errors is encoded using context coding.

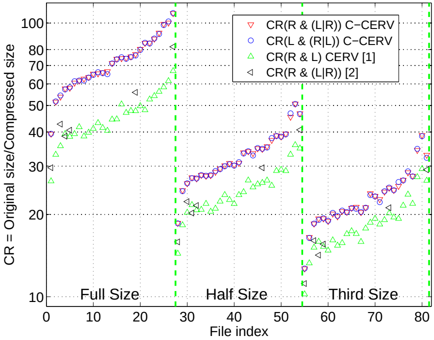


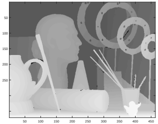


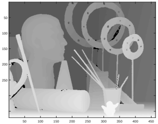

(a) Template $[\alpha_1, \ldots, \alpha_{16}]$ at $H(i,j)$ (b) Template $[\delta_0, \ldots, \delta_6]$ at $H^w(i,j)$

Figure: Crack-edges forming the context for encoding the crack-edge $H(i,j)$ of the left image $\mathbf{L}$: (a) The 16 causal neighbors from the crack-edge images $(H, V)$ extracted from the left image $\mathbf{L}$; (b) The 7 CEs unrestricted neighborhood centered at $H^w(i,j)$ from the crack-edge images $(H^w, V^w)$ extracted from the warped image $\mathbf{L}^w$.




Encode $D^L$ Encode $D^R$

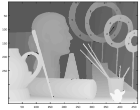
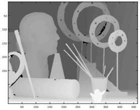
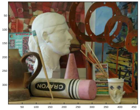
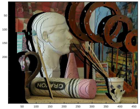
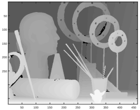


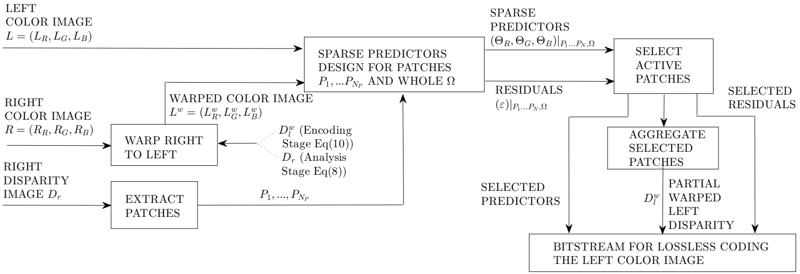


$\mathcal{L}(D^L|D^R)+$ $\mathcal{L}(D^R)$

Compression ratios when encoding the pairs ("disp1" and "disp5") using conditional coding for one image and CERV for the other, and finally when using only CERV for both images in the pair.

[2] I.Tabus, Patch-Based Conditional Context Coding of Stereo Disparity Images. IEEE Signal Processing Letters, 21:10, pp. 12201224, Oct. 2014

| | |
|---|---|
| C-CERV: Conditional encoding of DISPARITY maps | Encode $D^L$  $\mathcal{L}(D^L \vert D^R)+$ <br> Encode $D^R$  $\mathcal{L}(D^R)$ |
| SMPW: Conditional encoding of COLOR views using warping by disparity map | Encode $L$  $\mathcal{L}(L \vert D^R)+$ Compute $L^w$  <br> Encode $D^R$  $\mathcal{L}(D^R)$ Given $R$  |


LEFT
COLOR IMAGE
$L = (L_R, L_G, L_B)$
SPARSE PREDICTORS
DESIGN FOR PATCHES
$P_1, ...P_{N_P}$ AND WHOLE $\Omega$
SPARSE
PREDICTORS
$(\Theta_R, \Theta_G, \Theta_B)|_{P_1...P_N,\Omega}$
SELECT
ACTIVE
PATCHES
RESIDUALS
$(\varepsilon)|_{P_1...P_N,\Omega}$
SELECTED
RESIDUALS
RIGHT
COLOR IMAGE
$R = (R_R, R_G, R_B)$
WARPED COLOR IMAGE
$L^w = (L^w_R, L^w_G, L^w_B)$
WARP RIGHT
TO LEFT
$D^w_l$ (Encoding
Stage Eq(10))
$D_r$ (Analysis
Stage Eq(8))
AGGREGATE
SELECTED
PATCHES
RIGHT
DISPARITY
IMAGE $D_r$
EXTRACT
PATCHES
$P_1, ..., P_{N_P}$
SELECTED
PREDICTORS
$D^w_l$
PARTIAL
WARPED
LEFT
DISPARITY
BITSTREAM FOR LOSSLESS CODING
THE LEFT COLOR IMAGE

- The disparity image $D$ defines the partition into constant regions $\mathcal{P} = \{\Omega_1, \ldots, \Omega_N\}$.
- Overall description length (implementable codelength):

$$\mathcal{L}(L; \mathcal{P}, \Theta) = \mathcal{L}(\Gamma(\mathcal{P})) + \sum_{k=1}^{N} \mathcal{L}(D^w_{\Omega_k}) + \sum_{k=1}^{N} \mathcal{L}(\Theta(\Omega_k))$$

$$+ \sum_{k=1}^{N} \sum_{(i,j)\in\Omega_k} \mathcal{L}(\varepsilon^R(i,j)) + \mathcal{L}(\varepsilon^G(i,j)) + \mathcal{L}(\varepsilon^B(i,j)) \tag{1}$$

- $\mathcal{L}(\Gamma(\mathcal{P}))$ encoding the contours of the partition $\mathcal{P}$
- the term $\mathcal{L}(D^w_{\Omega_k})$ encodes the (constant) disparity value over the region $\Omega_k$
- $\mathcal{L}(\Theta(\Omega_k))$ accounts for encoding the predictor parameters (with the fractional part rounded to a precision of 16 bits)
- $\mathcal{L}(\varepsilon^R(i,j))$ encoded the residuals by using a context encoding scheme

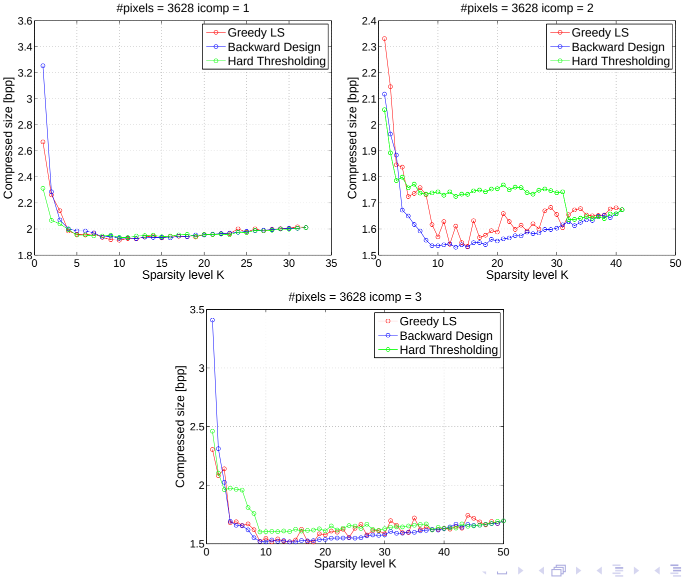
## Select Causal Prediction Mask

- Pick from WARPED image:
  - The NINE template centered at (i,j) from RED component
  - The NINE template centered at (i,j) from GREEN component
  - The NINE template centered at (i,j) from BLUE component
- Pick from the Left Image the Same Templates as in SMP
- Concatenate the two templates of regressors
- The predictors sizes are: 27+5, 27+14, and 27+23.
- Sparse design for all of them
- Alter the Template at Region Borders, to Ensure Region-contained Encoding
- Sparse Design By A Backward Truncation Policy

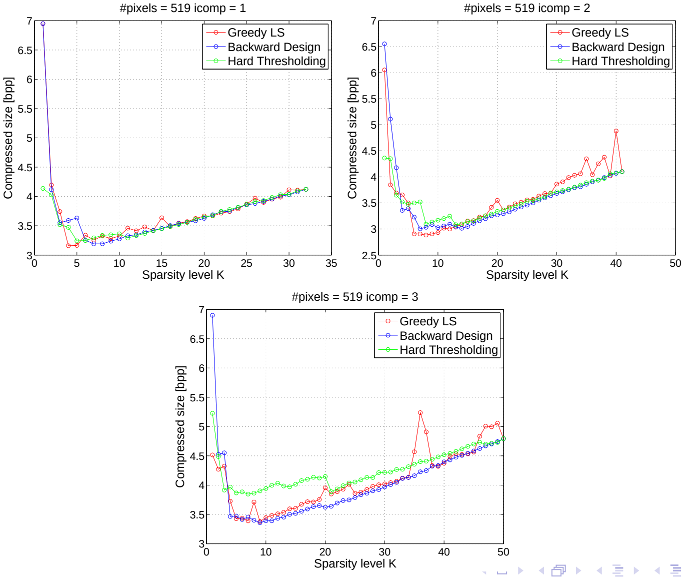
- Experiments with three methods for a predictor with sparsity $K$:
  - Greedy Least Squares - iterative, start from empty sparsity mask, add one entry at a time, high complexity for testing all candidates, good results
  - Hard Thresholding - non-iterative, order the regressors according to their scalar product with the free term vector and pick those with the largest $K$ scalar products
  - Backward design: compute the LS predictor for all regressors, and choose the $K$ regressors with largest absolute values of the LS-coefficients
- Backward design was used in the final experiments

- Typical predictive coding, similar to CALIC and LOCO-I
  - For each color component and at each pixel location $(i, j)$
  - Compute the prediction
  - Compute prediction residuals
  - Compute local gradients and determine the 16-valued context
  - Encode using arithmetic coding the prediction residual and maintain adaptively the distribution of residual at each context

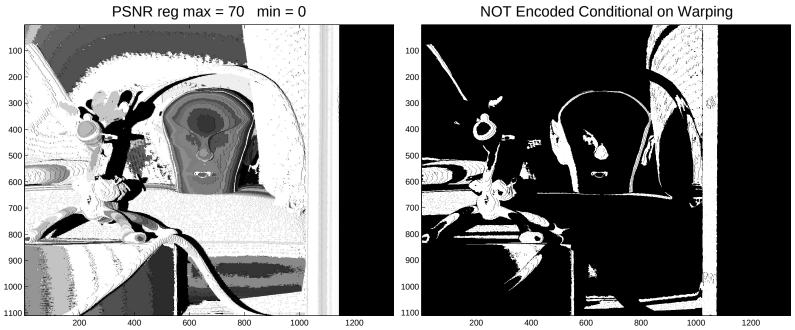
# Comparing the Codelength for 3 Sparse Design Methods (a region of 3600 pixels)



# Comparing the Codelength for 3 Sparse Design Methods (a region of 500 pixels)



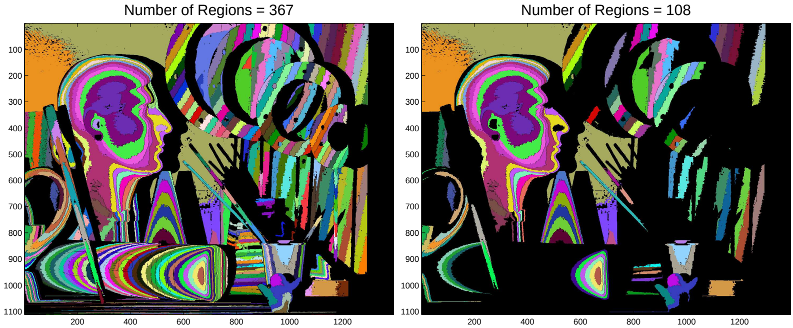
SPM = Sparse Predictive Models
SPMW = Sparse Predictive Models using Warped views
**Quality of regions: local PSNR and predictor competition**




(a) (b)

(a) PSNR over each region
(b) Not Encoded by Warping

# Constant disparity regions larger than the threshold at 500 pixels



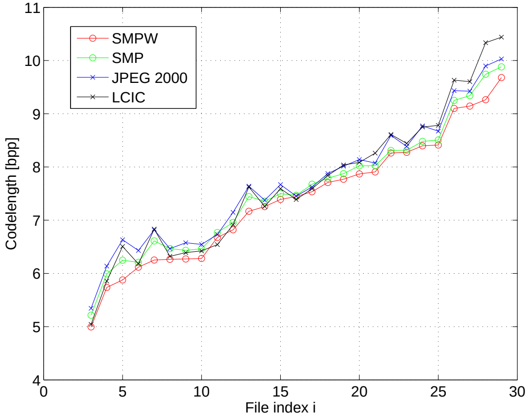

(a) (b)

(a) Regions in Warped Disparity larger than 500 pixels
(b) Regions in the Warped Disparity found worth of their own Sparse Predictor

| | SMPW | SMP | JPEG 2000 | LCIC[5] | CALIC | LOCO-I | PNG |
|---|---|---|---|---|---|---|---|
| Encoding **L** given **R** | **7.4021** | 7.5681 | 7.6999 | 7.6771 | 8.6605 | 9.169 | 10.395 |
| Encoding **R** given **L** | **7.3319** | 7.5005 | 7.6324 | 7.593 | 8.5175 | 9.0311 | 10.293 |

Average compression in bits per pixels (smaller is better) of the newly introduced sparse predictive methods, SMPW and SMP, and of several publicly available compression programs, over 27 images from Middlebury dataset. The second row shows results obtained when compressing the image **L** given **R** (as described in the text) while the last row shows results in the symmetric situation, when encoding **R** given **L**.

# Comparison with standard lossless compressors




Compressed size in bits per pixel (smaller is better) for the left color image **L** in the 27 files from Middlebury dataset using several publicly available lossless compressors and the newly presented methods. For the method SMPW only the right color image **R** was assumed available, the necessary parts of the disparity $D$ were encoded in the bitstream and accounted for in the compression size.

# Images sorted according to their SMPW sparse compressibility

4.81


5.63


5.8


6.07


6.07


6.15


6.2


6.29


6.68


6.77


7.08


7.23


7.3


7.4


7.43


7.59


7.61


7.65


7.87

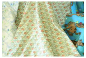

8.07

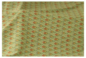

8.2


8.23


8.23


9.2


9.26


9.27


9.71


Applications:

1. Segmentation based on information theoretic region merging
   - Start from an oversegmentation of the image (e.g., constant patches)
   - Iteratively merge the regions of the image
     - Merge those two regions which result in the steepest Rate-Distortion move
     - In the resulting region use constant or planar model, based on MDL decision
2. Lossy coding: Encode each partial image using CERV

Possible applications of the models used in compression for image analysis

- Segmentation: region merging based on $\mathcal{L}(\text{SMPW})$
- Stereo matching based on $\mathcal{L}(\text{SMPW})$

- An algorithm for stereo matching estimates the disparity $D$ by minimizing the energy

$$
\begin{aligned}
E(D) &= \sum_{i,j} Score(L, R|D(i,j)) + \\
&+ \sum_{(i',j')\in\mathcal{N}(i,j)} \lambda_2 1_{|D(i',j')-D(i,j)|=1} \\
&+ \sum_{(i',j')\in\mathcal{N}(i,j)} \lambda_2 1_{|D(i',j')-D(i,j)|>1}
\end{aligned}
$$

 - $Score(L, R|D(i,j))$ is computed based on SSD of the pixels matched by $D(i,j)$
 - The smoothness of $D$ is enforced by the two terms weighted with $\lambda_1$ and $\lambda_2$.
 - The energy function can be minimized with efficient tools (e.g. graph cuts)
 - The final ranking of the algorithms is based on the percentage of erroneous pixels, where $|D(i',j') - D(i,j)| > 1$.

- The disparity image $D$ defines the partition into constant regions $\mathcal{P} = \{\Omega_1, \ldots, \Omega_N\}$.
- Overall description length (implementable codelength):

$$\mathcal{L}(L; \mathcal{P}, \Theta) = \mathcal{L}(\Gamma(\mathcal{P})) + \sum_{k=1}^{N} \mathcal{L}(D^w_{\Omega_k}) + \sum_{k=1}^{N} \mathcal{L}(\Theta(\Omega_k))$$
$$+ \sum_{k=1}^{N} \sum_{(i,j) \in \Omega_k} \mathcal{L}(\varepsilon^R(i,j)) + \mathcal{L}(\varepsilon^G(i,j)) + \mathcal{L}(\varepsilon^B(i,j)) \quad (2)$$

- $\mathcal{L}(\Gamma(\mathcal{P}))$ encoding the contours of the partition $\mathcal{P}$ (enforces SMOOTHNESS of $D$)
- the term $\mathcal{L}(D^w_{\Omega_k})$ encodes the (constant) disparity value over the region $\Omega_k$ (MODEL cost)
- $\mathcal{L}(\Theta(\Omega_k))$ accounts for encoding the predictor parameters (with the fractional part rounded to a precision of 16 bits) (MODEL cost)
- $\mathcal{L}(\varepsilon^R(i,j))$ encodedthe residuals by using a context encoding scheme (MATCHING cost)

- Depth or disparity maps are much more compressible than natural grayscale or images. Compression ratios: 50 to 200
- The most efficient representation of depth map is into constant patches, to be encoded by their contours
- Exploiting the redundancy Left-Right leads to high gains in lossless compression of depth images
- The tools for lossless encoding include: context trees for the crack-edges forming the contour and list encoding of depth values in a graph of neighbor patches
- Left-Right redundancy brings considerable savings in bits in lossless compression of color views and *sparse predictive coding* produces competitive lossless compression results.
- The combined cost $CL(C_L|C_R) + CL(D_R)$ is better than $CL(C_L)$, which was not true for lossy compression in [1] (there, only $CL(C_L|C_R) \leq CL(C_L)$ was achieved).
- Additionally, the disparity image (encoded with CERV coder) becomes available to the decoder for free.

# References


M. Zamarin, P. Zanuttigh, S. Milani, G. M. Cortelazzo, and S. Forchhammer,
"A joint multi-view plus depth image coding scheme based on 3D-warping,"
in *3DVP, Firenze, Italy*, 2010.

K. Müller, H. Schwarz, D. Marpe, C. Bartnik, S. Bosse, H. Brust, T. Hinz, H. Lakshman, P. Merkle, F. H. Rhee, G. Tech, M. Winken, and T. Wiegand,
"3D high-efficiency video coding for multi-view video and depth data,"
*IEEE Transactions on Image Processing*, vol. 22, pp. 3366–3378, 2013.

M. Domanski, O. Stankiewicz, K. Wegner, M. Kurc, J. Konieczny, J. Siast, J. Stankowski, R. Ratajczak, and T. Grajek,
"High efficiency 3D video coding using new tools based on view synthesis,"
*IEEE Transactions on Image Processing*, vol. 22, pp. 3517–3527, 2013.

I. Tabus, I. Schiopu, and J. Astola,
"Context coding of depth map images under the piecewise-constant image model representation,"
*IEEE Transactions on Image Processing*, vol. 22, pp. 4195–4210, November 2013.

Seyun Kim and Nam Ik Cho,
"Hierarchical prediction and context adaptive coding for lossless color image compression,"
*IEEE Transactions on Image Processing*, vol. 23, no. 1, pp. 445–449, Jan 2014.

I. Tabus and P. Astola,
"Sparse prediction for compression of stereo color images conditional on constant disparity patches,"
*in Proceedings of 3DTV-Con 2014*, Budapest, July 2014, pp. 14.

I. Tabus,
"Patch-Based Conditional Context Coding of Stereo Disparity Images,"
*IEEE Signal Processing Letters*, vol. 21, no.10, pp.1220,1224, Oct. 2014.