Layer based multiview image compression

Andriy Gelman
Supervisors: Dr. P.L. Dragotti, Dr. V. Velisavljević

Imperial College London

October 7, 2009

Presentation Outline

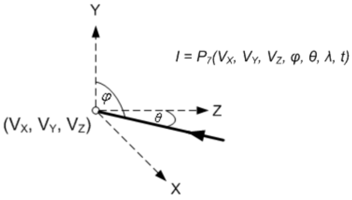
Presentation Outline

Image Based Rendering - IBR

- Multiview images model a 3D representation of a scene
- IBR is defined as estimating a continuous high dimensional function which stores all of the visual information from its samples
- Advantages - Photo-realistic results, immersive viewing. Important for 3D and Free viewpoint TV

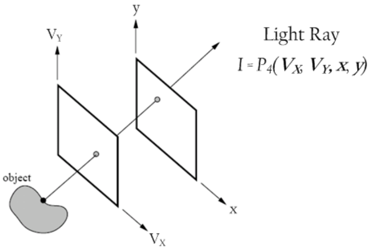
IBR - Plenoptic Function

- 7D function stores all of the visual information.
- Assume that the space is filed with infinitesimally thin rays of light.
- Each light ray can be uniquely parameterized by a 3D point in space $(V_x, V_y, V_z,)$, DOA (θ, ϕ), time t, and wavelength λ.
- Simplifications - Analyze static, mono-chromatic scenes. Assume intensity of light does not change along the path. Reduce size by restricting the field of view.



Light Field Rendering

- Assumptions - Static and monochromatic scene. Intensity of a light ray does not change along its path.
- Parameterize the light rays by their intersection with two planes.
- Light field can be modeled by using a 2D array of images.
- Disadvantage - A large number of images is required to achieve artifact-free rendering. Typical uncompressed light field $\approx$ 1GB



Multiview image redundancy

- Multiview images are highly redundant - inter and intra frame correlation.
- We expect to achieve high compression rates.
- Data structure is linear, inversely proportional to the depth.

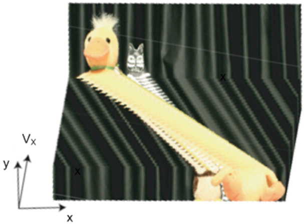
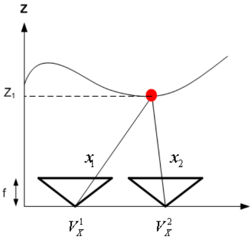
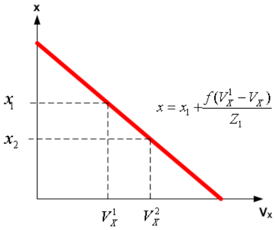


Figure: Multiview image cross-section [Berent08]

Multiview image structure

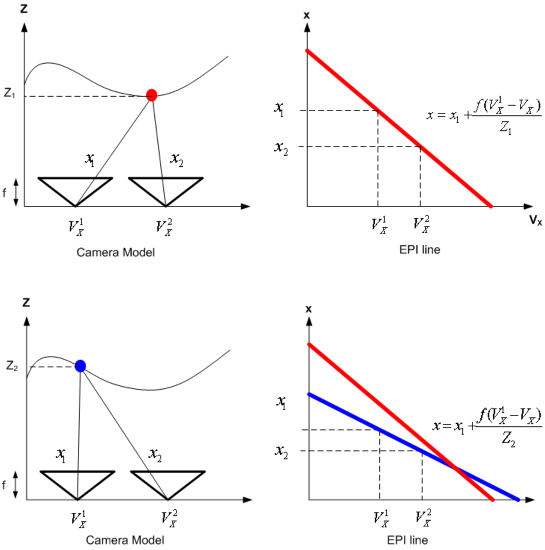


Camera Model



EPI line

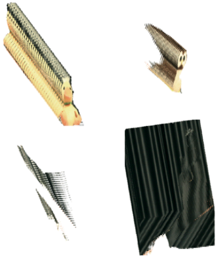
Multiview image structure



Presentation Outline

Layer based representation

- Aim - Partition data into coherent layers. Exploit the data redundancy in each layer in the compression stage.
- A coherent layer should be highly correlated in the view dimension, and contain pixels belonging to the same object or objects at the same depth.
- Introduce data segmentation tools and apply them to layer extraction in the multi-view image data set.



Region based data segmentation

- Aim - Partition data into homogenous regions. E.g. Minimize statistical features such as the variance.
- Mathematical formulation - Partition $\mathcal{D}(\mathbf{x}) \subset \mathbb{R}^m$ into subsets $\mathcal{H}(\mathbf{x})$ and $\overline{\mathcal{H}}(\mathbf{x})$ where $\mathbf{x} \epsilon \mathbb{R}^m$
- Boundary separating the two regions is denoted by $\Gamma(\boldsymbol{\sigma}) \subset \mathbb{R}^m$
- The boundary is closed and can be parameterized using one less variable than the dimension of the data structure.

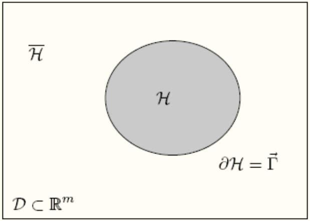


Figure: [Berent08]

Region based data segmentation - 2

- Set the complete problem in an optimization framework. The function is minimized when the correct segmentation is obtained.

$$\Gamma = \arg\min\{J(\Gamma)\},$$

$$J(\Gamma) = \int_{\mathcal{H}} d(\mathbf{x}, \mathcal{H})\, d\mathbf{x} + \int_{\overline{\mathcal{H}}} d\left(\mathbf{x}, \overline{\mathcal{H}}\right) d\mathbf{x} + \int_{\Gamma} \mu d\boldsymbol{\sigma}.$$

- The descriptors d, are designed to measure the homogeneity of each region, and are therefore region dependent.
- Descriptors are chosen according to the statistical properties we wish to minimize, e.g. variance.
- The cost function also contains an additional regularization term, which acts to minimize the length of the boundary.

Dynamic boundary evolution

- Introduce a dynamical scheme and evolve the boundary towards the ideal segmentation.
- Model the boundary evolution using a P.D.E -

$$\frac{\partial \Gamma(\boldsymbol{\sigma}, \tau)}{\partial \tau} = v(\boldsymbol{\sigma}, \tau) = F(\boldsymbol{\sigma}, \tau)\, n_\Gamma(\boldsymbol{\sigma}, \tau),$$

- Velocity vector v determines how the boundary evolves. Modeled by a force F, acting in the outward normal direction.
- Determine velocity vector such that boundary evolves towards the ideal segmentation.

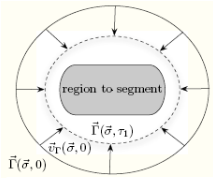


Figure: Active contour evolution [Berent08].

Dynamic boundary evolution

- Obtain a relationship between velocity v, and the cost function.

$$\frac{\partial J(\Gamma(\tau))}{\partial \tau} = \int_{\Gamma(\tau)} \left[d(\mathbf{x}, \mathcal{H}) - d\left(\mathbf{x}, \overline{\mathcal{H}}\right) + \mu k\right] (v_\Gamma . n_\Gamma)\, d\sigma,$$

- Choose a velocity vector to ensure that the cost function reduces with each iteration.

$$v = \left[d\left(\mathbf{x}, \overline{\mathcal{H}}\right) - d(\mathbf{x}, \mathcal{H}) - \mu k\right] (n_\Gamma).$$

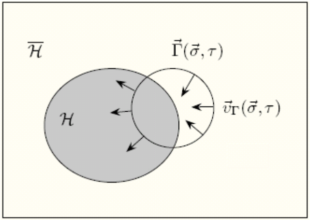


Figure: Region competition [Berent08]

Multiview image layer extraction - Problem formulation

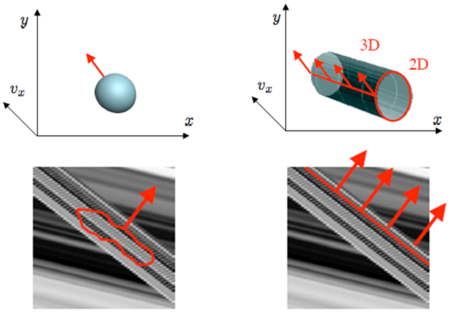
- Multiview data is a 1D array of images - $\mathcal{D} \subset \mathbb{R}^3$, $\mathbf{x} = [x, y, v_x] \ \epsilon \ \mathbb{R}^3$
- N unknown layer boundaries and the associated disparities p.

$$J(\Gamma_1, \Gamma_2, \ldots, \Gamma_N, p_1, p_2, \ldots, p_N) = \sum_{i=1}^{N} \int\int\int_{\mathcal{H}_i^{\perp}} d(\mathbf{x}, p_i)\, d\mathbf{x},$$

- Descriptor is chosen to minimize the variance along the EPI lines, $d(\mathbf{x}, p_i) = [I(\mathbf{x}) - m(\mathbf{x}, p_n)]^2$
- $m(\mathbf{x}, p_n)$, is the mean of the EPI line which passes through point $\mathbf{x}$ and has a disparity p_n.
- Evolve one boundary at time, assume that the disparities are known.

Multiview image layer extraction - Constrained velocity vector

- Velocity vector is 3D - Spatial consistency cannot be guaranteed
- The evolution can be constrained to a 2D vector [Berent08]



3D active contour

Constrained active contour [Berent08]

Multiview image layer extraction - Algorithm overview

Algorithm Outline

1. Initialize the layer
2. Estimate the disparities - $p_1, p_2, \ldots, p_N$
3. For each layer compute the velocity vector and evolve the boundary

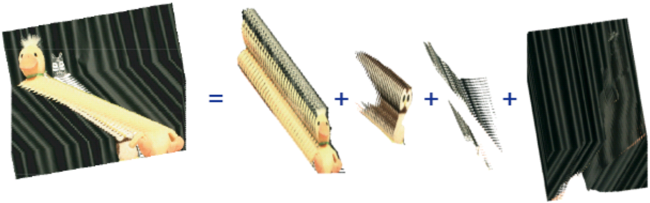
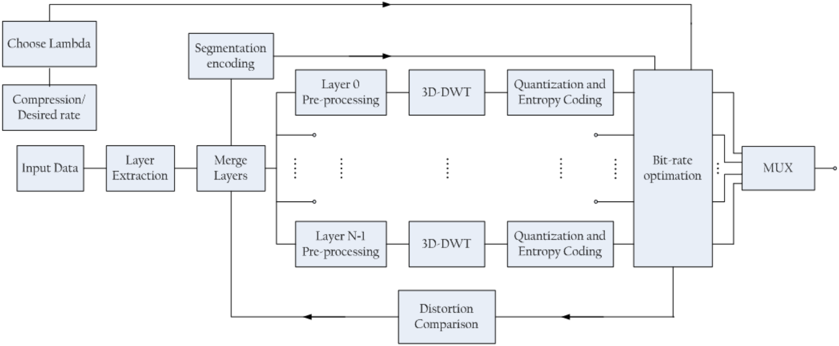


Figure: Extracted layers [Berent08]

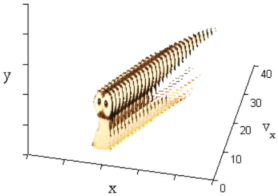
Presentation Outline

Layer based multiview image compression

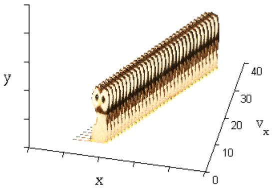


Layer pre-processing

- Algorithm ensures the data is spatially consistent in the view dimension
- Occluded regions are explicitly defined by the extraction algorithm
- Missing data is filled by the average value along the EPI lines



(a) Original owl layer

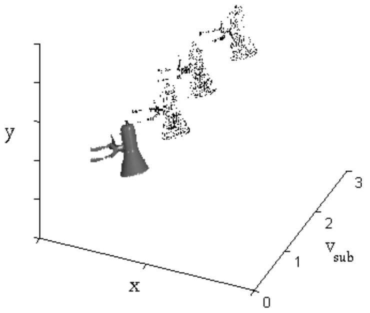


(b) Owl layer with pre-processing

1D Disparity Compensated DWT

- De-correlate the layers along the view dimension
- Use a lifting implementation of the DWT to reduce computational complexity and ensure the transform is invertible
- Disparity compensation is incorporated into the transform
- Multi-resolution representation can be achieved by applying the DWT on the low-pass component

$$\mathcal{L}_e[n] = \frac{P_e[n] - \mathcal{W}\{P_o[n]\}}{2}$$
$$\mathcal{L}_o[n] = P_o[n] + \mathcal{W}\{\mathcal{L}_e[n]\}$$

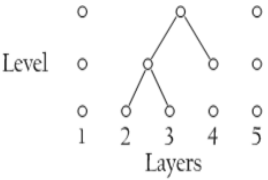


2D Shape-adaptive transform DWT and Entropy Coding

- Data is de-correlated in the spatial dimension using a Shape Adaptive 2D-DWT.
- Segmentation is known and and we use symmetric extensions to reduce the boundary effects
- Symmetric extensions lead to symmetric wavelet coefficients
- Number of decompositions is determined by the spatial size of the layers
- The data is partitioned into 2D blocks and each one is entropy coded using context adaptive arithmetic coding
- Output rate is controlled using a Lagrangian multiplier λ. Use operational rate-distortion curves to allocate an optimum number of bits to each block.

Layer-based representation bit-allocation

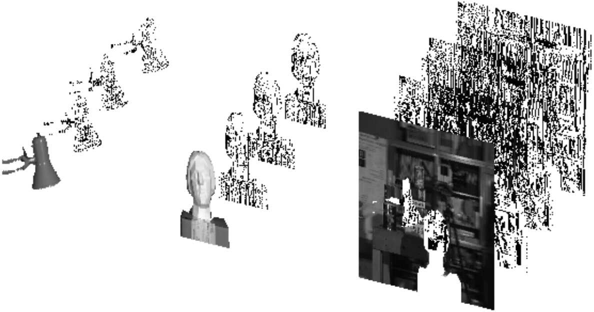
- How many bits should be allocated to the representation? Must be correct in the rate-distortion sense.
- A layer - based representation with a larger number of layers leads to better disparity compensation, however requires more bits.
- Propose a bottom-up greedy approach.
- Merge two layers if: $D_M + \lambda R_M < (D_{l_1} + D_{l_2}) + \lambda (R_{l_1} + R_{l_2})$



1. Initialize N layers and choose λ
2. While merging reduces the Lagrangian cost
    3. Choose two layers with the smallest disparity
    4. Merge Layers

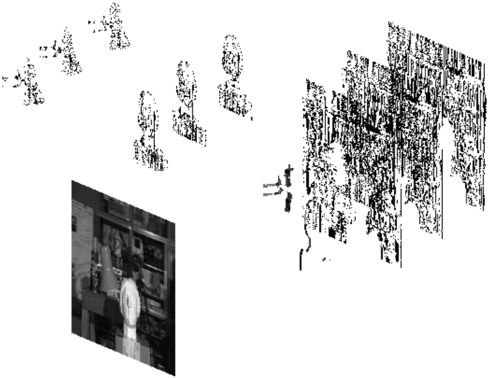
Algorithm 2 - Low pass recombined

- A complicated scene requires a large number of layers.
- A small number of spatial decompositions can be applied on each layer.
- Inter-layer correlation is not exploited



Algorithm 2 - Low pass recombined

- Recombine low-pass subband components prior to applying the spatial transform
- Layers are not orthogonal - additional pixels have to transmitted



Presentation Outline

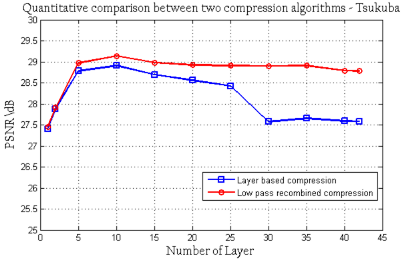
Analysis Data Sets

Animal Farm – 16×235×625
[Berent08]

Tsukuba – 4×284×382
[Middlebury02]

Teddy – 8×375×411
[Middlebury02]

Algorithm comparison - Independent and Low pass recombined

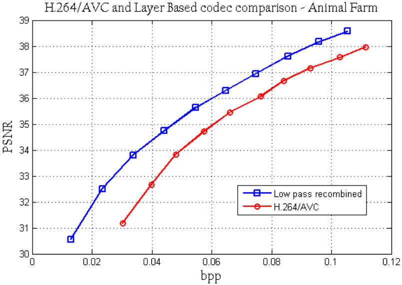


Encoding bit rate = 0.105bpp

(a) Algorithm 2 - Low Pass Recombined, PSNR = 28.91dB - 30 layers

(b) Algorithm 1 - Layer Based Compression, PSNR = 27.65dB - 30 layers

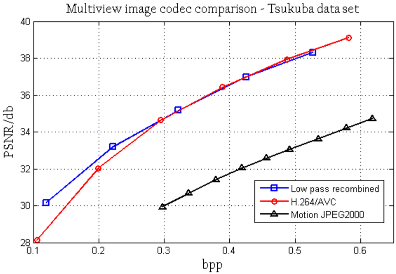
Algorithm comparison - Animal farm



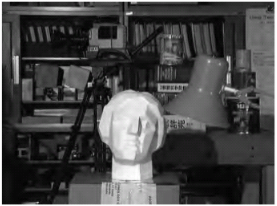
(a) H.264/AVC - PSNR = 31.1839dB at 0.0303bpp

(b) Low pass recombined - PSNR = 33.5362dB at 0.0316bpp

Algorithm comparison - Tsukuba

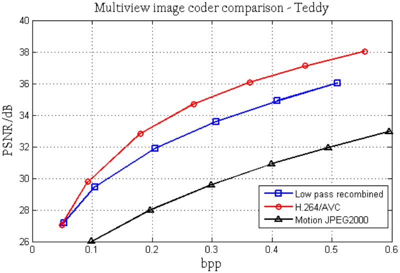


(a) JPEG2000 - PSNR = 27.83dB at 0.197bpp



(b) Low pass recombined - PSNR = 32.1354dB at 0.183bpp

Algorithm comparison - Teddy



(a) JPEG2000 - PSNR = 27.9986dB at 0.1967bpp

(b) Low pass recombined - PSNR = 31.91dB at 0.2055bpp

Presentation Outline

Conclusion

- Presented structure of multiview data
- Analyzed the coherent layer extraction algorithm
- Presented two layer based multiview image compression algorithms
- Our layer based compression algorithm outperforms JPEG2000 on all data sets and H.264/AVC on two out of three data sets.

Future Work

- The multiview compression algorithm can be extended to operate on a Light Field or an additional time dimension. The layer extraction algorithm scales naturally to any number of dimensions.
- A number of features such as random access, scalability and complexity need to be addressed. Distributed Source Coding (DSC) theory may be used to find the optimal trade-off while analyzing these conflicting features.
- As shown in the results, when the 3D scene is complicated the layer based representation does not capture the full complexity of the scene. This issue should be addressed to ensure our compression algorithms are competitive when encoding natural data.

Thank you for listening!

Multiview image layer extraction - constrained velocity vector

- Express cost function in the same format as the segmentation of two regions.

$$J\left(\Gamma_l\left(\tau\right)\right) = \int\int\int_{\mathcal{H}_l(\tau)^\perp} d\left(\mathbf{x}, p_l\right) d\mathbf{x} + \int\int\int_{\overline{\mathcal{H}_l(\tau)^\perp}} d^{out}\left(\mathbf{x}, p_1, p_2, \ldots, p_N\right) d\mathbf{x},$$

- Second integral is the combination of all the other region boundaries.
- Obtain the boundary velocity vector by differentiating the cost function w.r.t. τ

$$v_l = \left[d^{out}\left(\mathbf{x}, p_1, p_2, \ldots, p_N\right) - d\left(\mathbf{x}, p_l\right)\right] n_\Gamma.$$

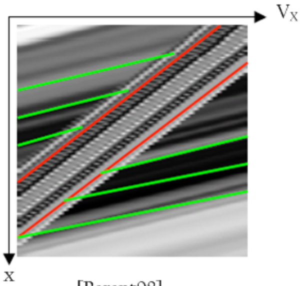
- Velocity vector evolves the complete 3D layer. Use spatial consistency to simplify the problem to the evolution of a 2D layer.

$$\gamma_l = \left[\int d^{out}\left(\mathbf{x}, p_1, p_2, \ldots, p_N\right) O_l\left(\mathbf{x}\right) dv_x - \int d\left(\mathbf{x}, p_l\right) O_l\left(\mathbf{x}\right) dv_x\right] n_\gamma,$$

Velocity vector analysis

- Velocity vector - competition between variance of the EPI lines.

$$\gamma_l = \left[\int d^{out}\left(\mathbf{x}, p_1, p_2, \ldots, p_N\right) O_l\left(\mathbf{x}\right) dv_x - \int d\left(\mathbf{x}, p_l\right) O_l\left(\mathbf{x}\right) dv_x \right] n_\gamma,$$



[Berent08]

Algorithm Overview:
1. Initialize the layers
2. Estimate disparities $p_1, p_2, \ldots, p_N$
3. for each layer{
compute velocity vector v
evolve the boundary}

Velocity vector analysis

- Velocity acts to minimize the cost function.
- Region competition.
- Boundary points belonging to the region will create a positive force and will be evolve the boundary to incorporate the point and vice versa.

$$v = \left[d\left(\mathbf{x}, \overline{\mathcal{H}}\right) - d\left(\mathbf{x}, \mathcal{H}\right) - \mu k\right]\left(n_\Gamma\right).$$

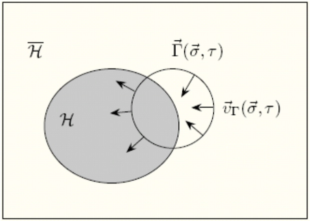


Figure: Region competition [Berent08]