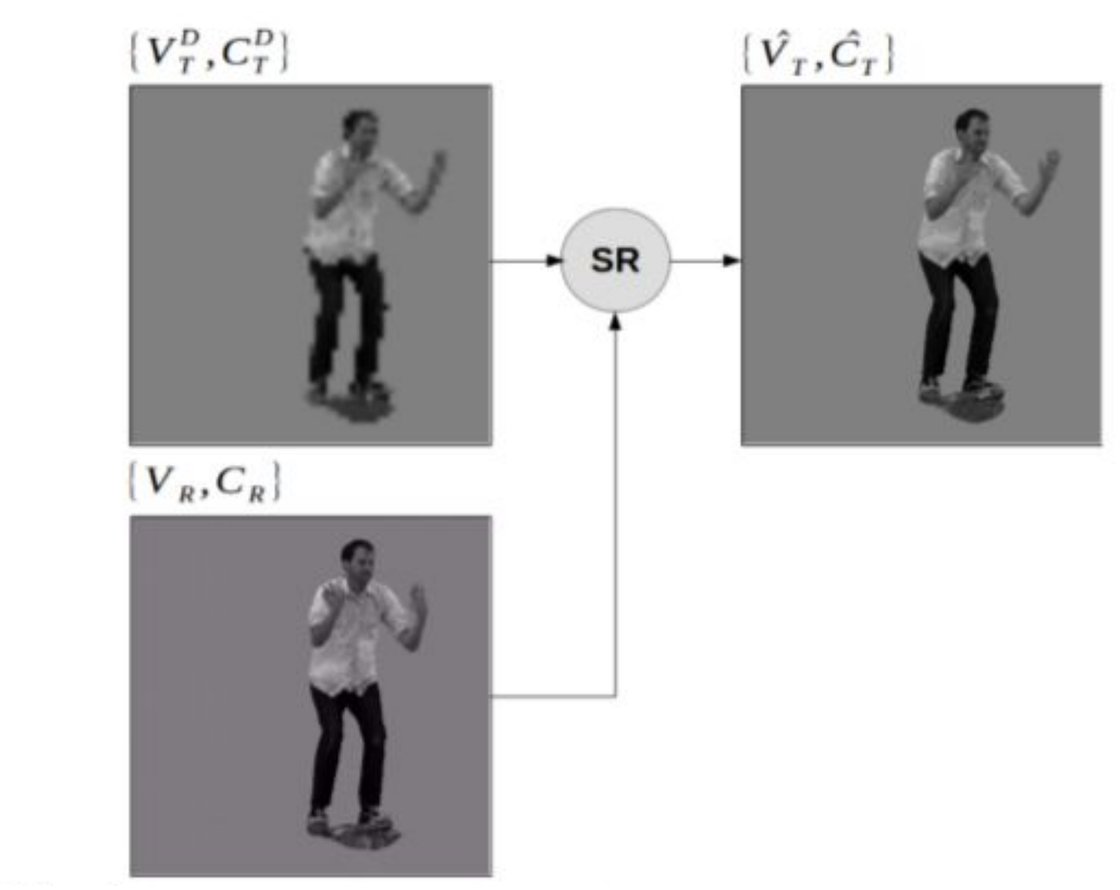
EXAMPLE-BASED SUPER-RESOLUTION FOR POINT-CLOUD VIDEO

Diogo C. Garcia, Tiago A. Fonseca and Ricardo L. de Queiroz
Universidade de Brasilia, Brasilia, Brasil
diogogarcia@unb.br, tiago@image.unb.br and queiroz@ieee.org

ABSTRACT

- Example-based super-resolution (SR) framework for point clouds (PCs):
  - Super-resolve low-resolution frames based on similarities with adjacent full-resolution frames.
- Several processing tools can be derived: compression, denoising and error concealment.
- Results: average gain of 1.18 dB over low-pass versions of the point-cloud, for a projection-based distortion metric.

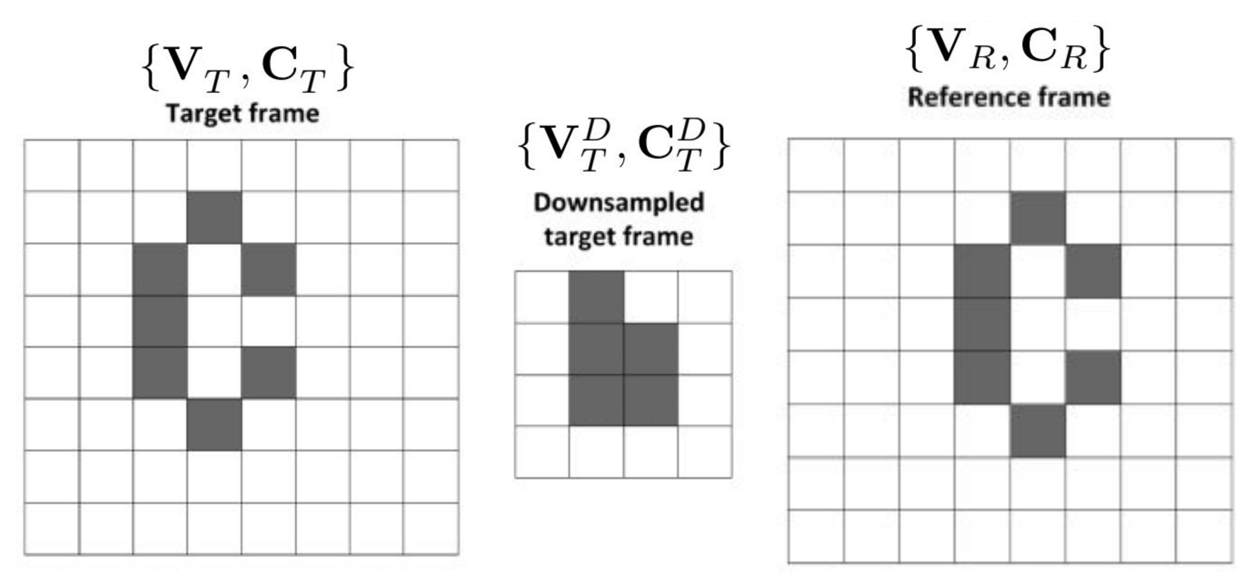


INTRODUCTION

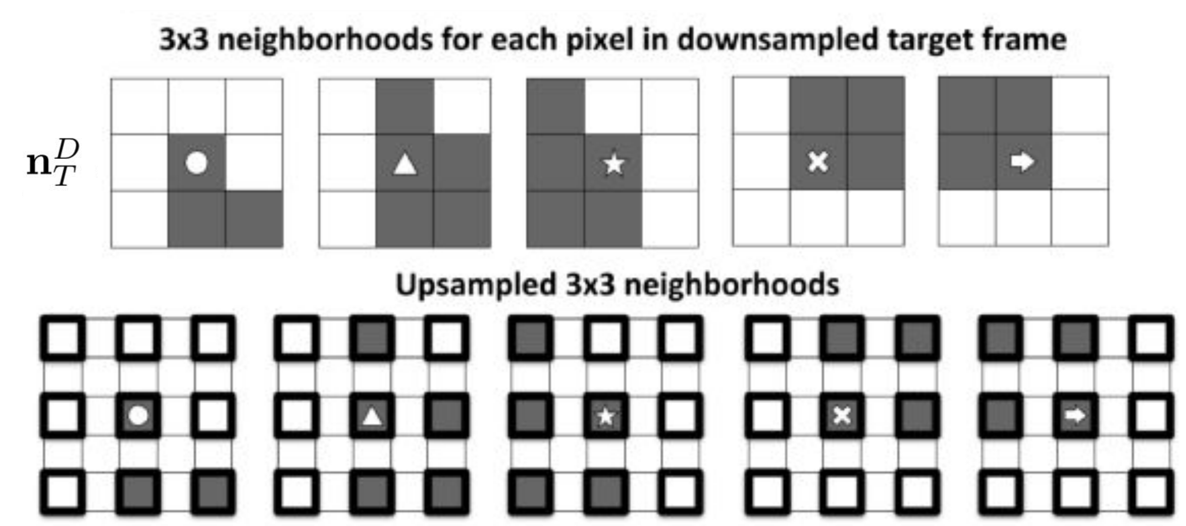
- AR/VR applications ⇒ greater interest in 3D signals (capturing, processing and rendering).
- No established standards.
- We focus on signals captured using a set of RGBD cameras: *voxelized point clouds.*
- Geometry representation: octrees.
  - Data compression
  - Fast search
  - Spatial scalability and mixed-resolution scenarios.
- SR for 3D signals: depth-map resolution increase.
- Our goal is to <u>infer high-frequency content from time-adjacent frames</u>.

PROPOSED METHOD

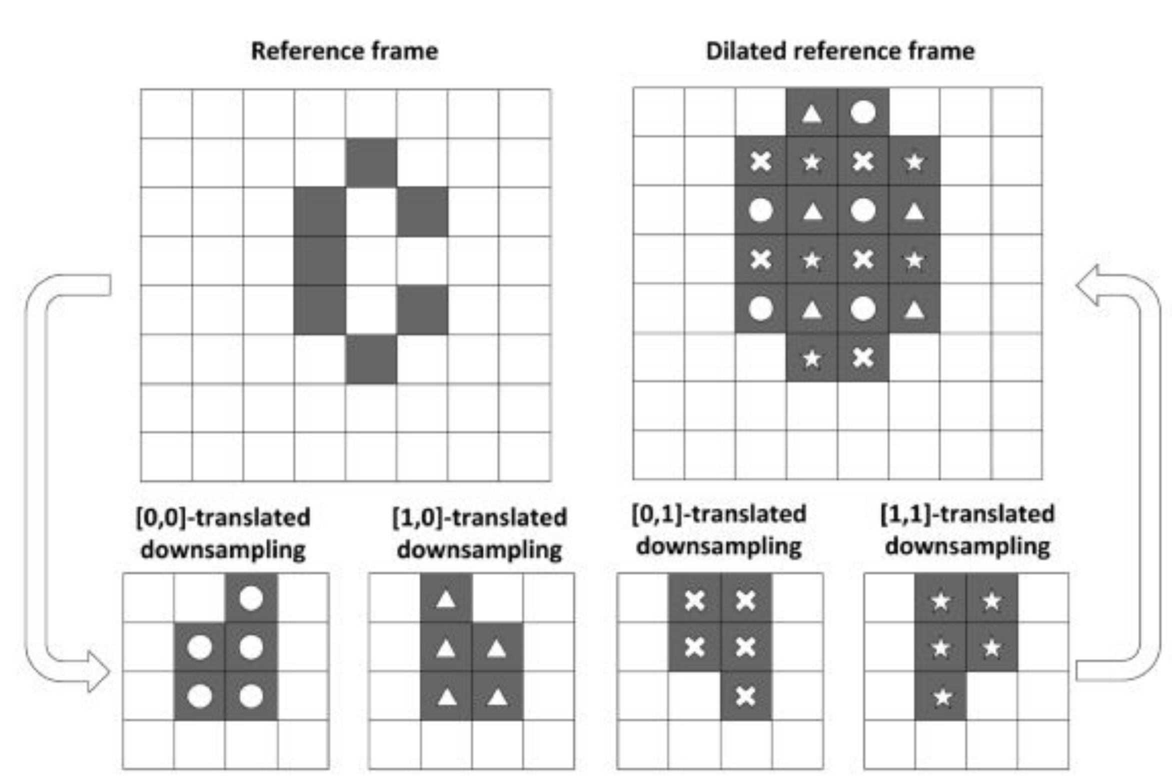
- Super-resolve a low-resolution point cloud with high-resolution information from an adjacent frame.



- For each voxel in $\mathbf{V}_T^D$, D_F x D_F x D_F voxels should be inferred, where D_F is the downsampling factor.
- Extract high-resolution from reference using motion estimation between $\{\mathbf{V}_T^D, \mathbf{C}_T^D\}$ and $\{\mathbf{V}_R^D, \mathbf{C}_R^D\}$ based on an *N x N x N* neighbourhood around each voxel.



- To be able to find 1-voxel motion at full resolution, we **dilate** the downsampled reference signal.



- Considering a search window, we minimize the cost function

$$C(i,j) = \frac{H(i,j) + wD(i,j)}{w+1},$$

where

- $H(i,j)$ is the hamming distance between $\mathbf{n}_T^D(i)$ and $\mathbf{n}_R^{DL}$;
- $D(i,j)$ is the Euclidean distance between $\mathbf{V}_T^D(i)$ and $\mathbf{V}_R^{DL}(j)$;
- w is the inverse of the Euclidean distance between the centers of mass of $\mathbf{V}_T^D$ and $\mathbf{V}_R^{DL}$.

METRICS

- Projection-based metric (PPSNR)
  - Orthographic projections on the surrounding cube faces: point cloud converted to 2D images.
  - Evaluate PSNR between the original and the super-resolved projections.

- Geometric distortion metric (GPSNR)
  - Point-to-plane distances between original and super-resolved pointclouds.

RESULTS

- Super-resolved PCs compared to a low-pass version (upsampled version of a downsampled reference PC).

Table 2. SR performance improvements. PPSNR Gains and GPSNR Gains stand for the average gains in projected PSNR and in geometric quality metric [14, 2], respectively. All values are in dB.

Sequence	PPSNR Gains	GPSNR Gains
Andrew	0.76	4.99
David	1.01	4.25
Loot	1.84	5.40
Man	1.93	∞
Phil	0.27	4.61
Ricardo	1.24	5.16
Sarah	1.18	4.64
Average	**1.18**	**4.84**

- Subjective evaluations for the best (Man) SR and the worst (Phil) SR performance are allowed by views comparison.

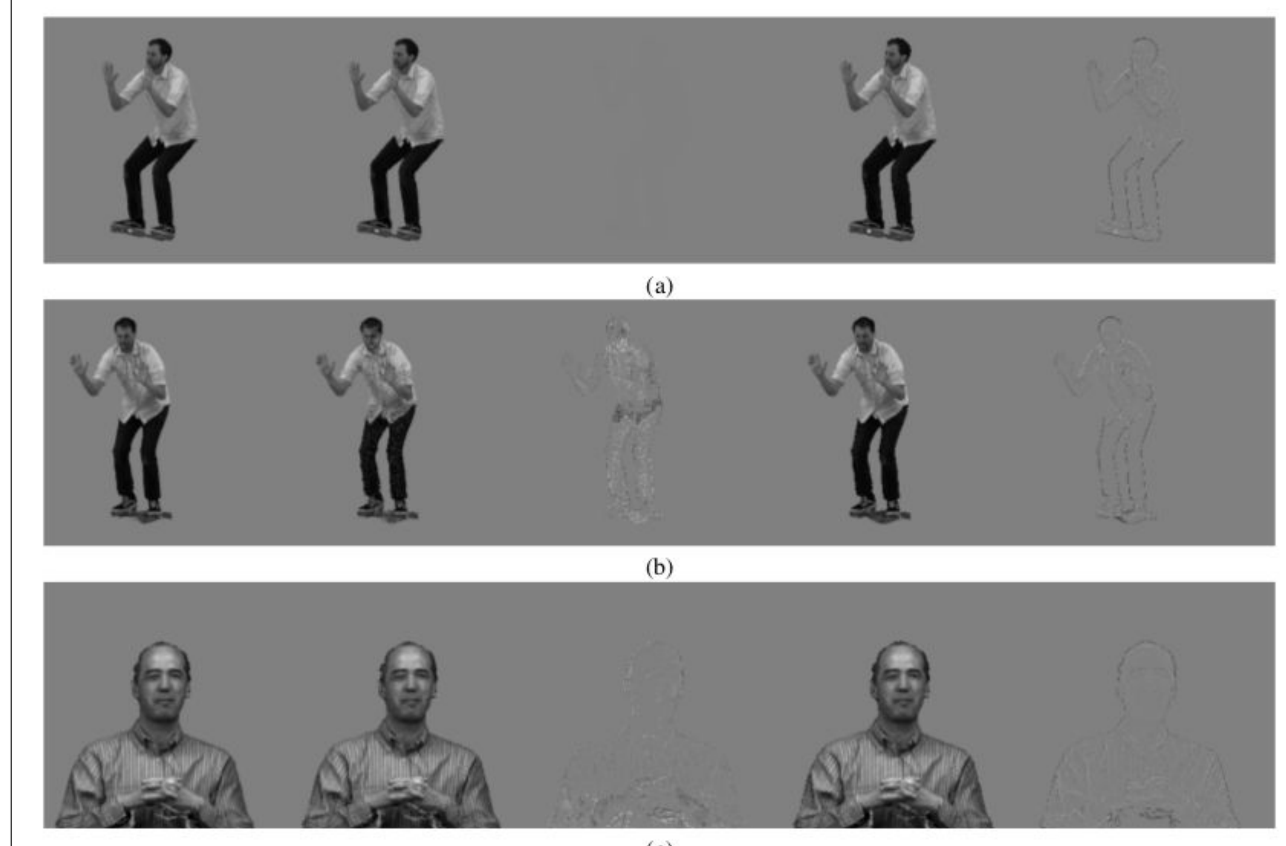


Fig. 5. Point-cloud projections for sequences (a)-(b) *Man*, frames 23 and 93, and (c) *Phil*, frame 175. For each image, from left to right, the columns correspond to the projections of: the original signal, the super-resolved signal, the residue of the super-resolved signal, the low-pass signal and the residue of the low-pass signal.

CONCLUSIONS

- Framework <u>successfully inferred high-frequency</u> by exploring similarities between adjacent point-cloud frames.
- Results can benefit a point-cloud encoding framework for efficient transmission, error concealment and storage.

ACKNOWLEDGEMENT

- Work partially supported by CNPq under grant 308150/2014-7.