46th International Conference on Acoustics, Speech, and Signal Processing (ICASSP), June 6th-11th 2021

Raw Data Processing for Practical Time-of-Flight Super-resolution

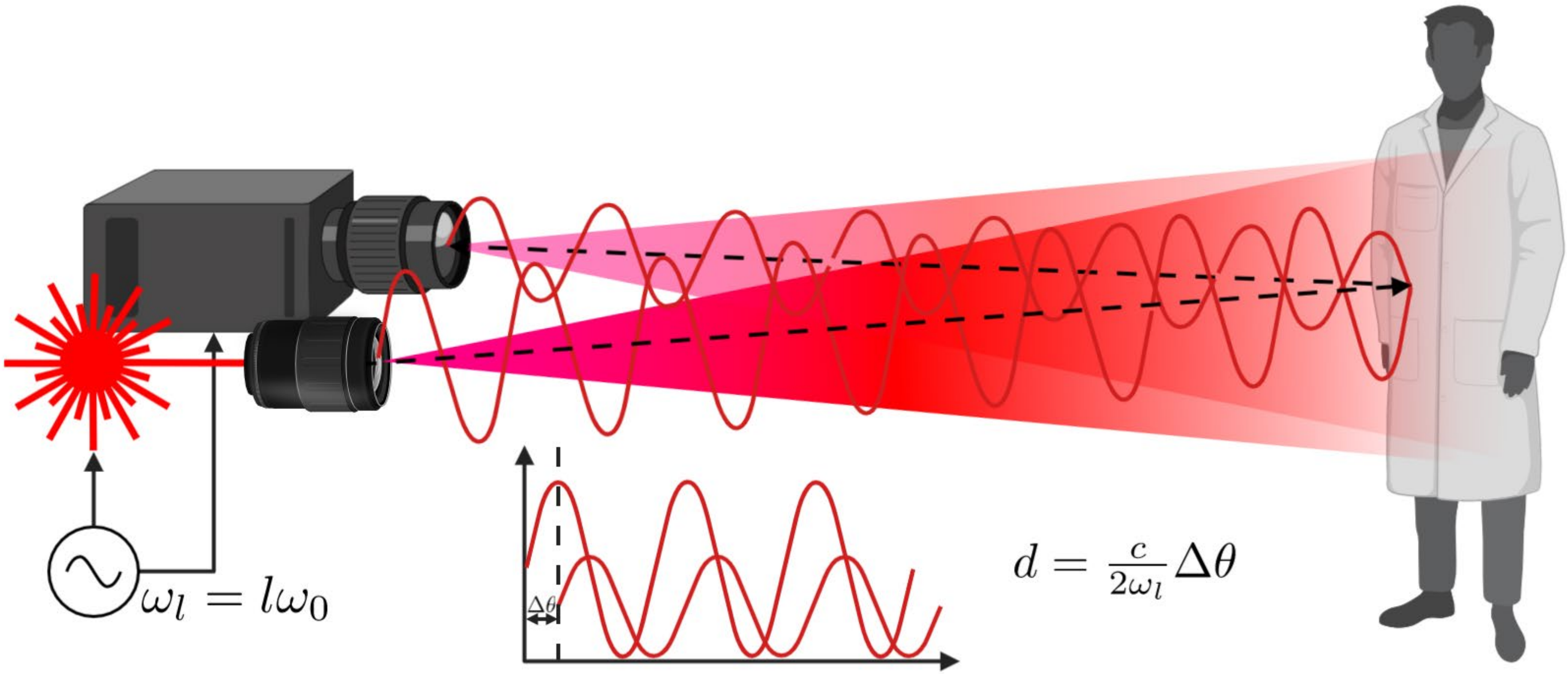
Miguel Heredia Conde
Center for Sensorsystems, University of Siegen
Paul-Bonatz-Str.9-11,57076 Siegen, Germany

Outline

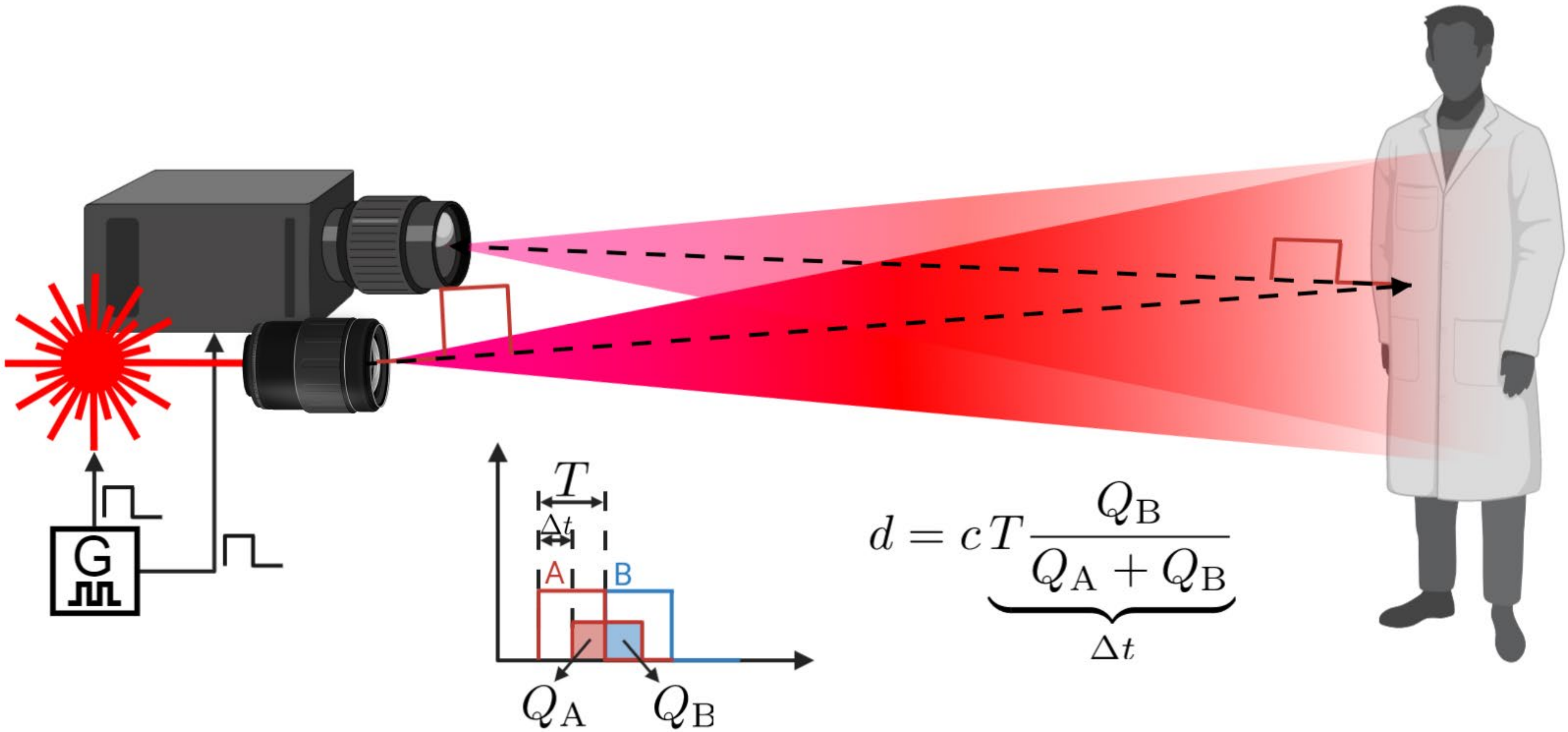
1. Introduction
2. Why ToF Super-resolution?
3. Methods for ToF Super-resolution
4. Intra-frame ToF Super-resolution from Raw Data
5. Experimental Results
6. Conclusion

1. Introduction

Continuous Wave Time-of-Flight Imaging



Pulsed Time-of-Flight Imaging



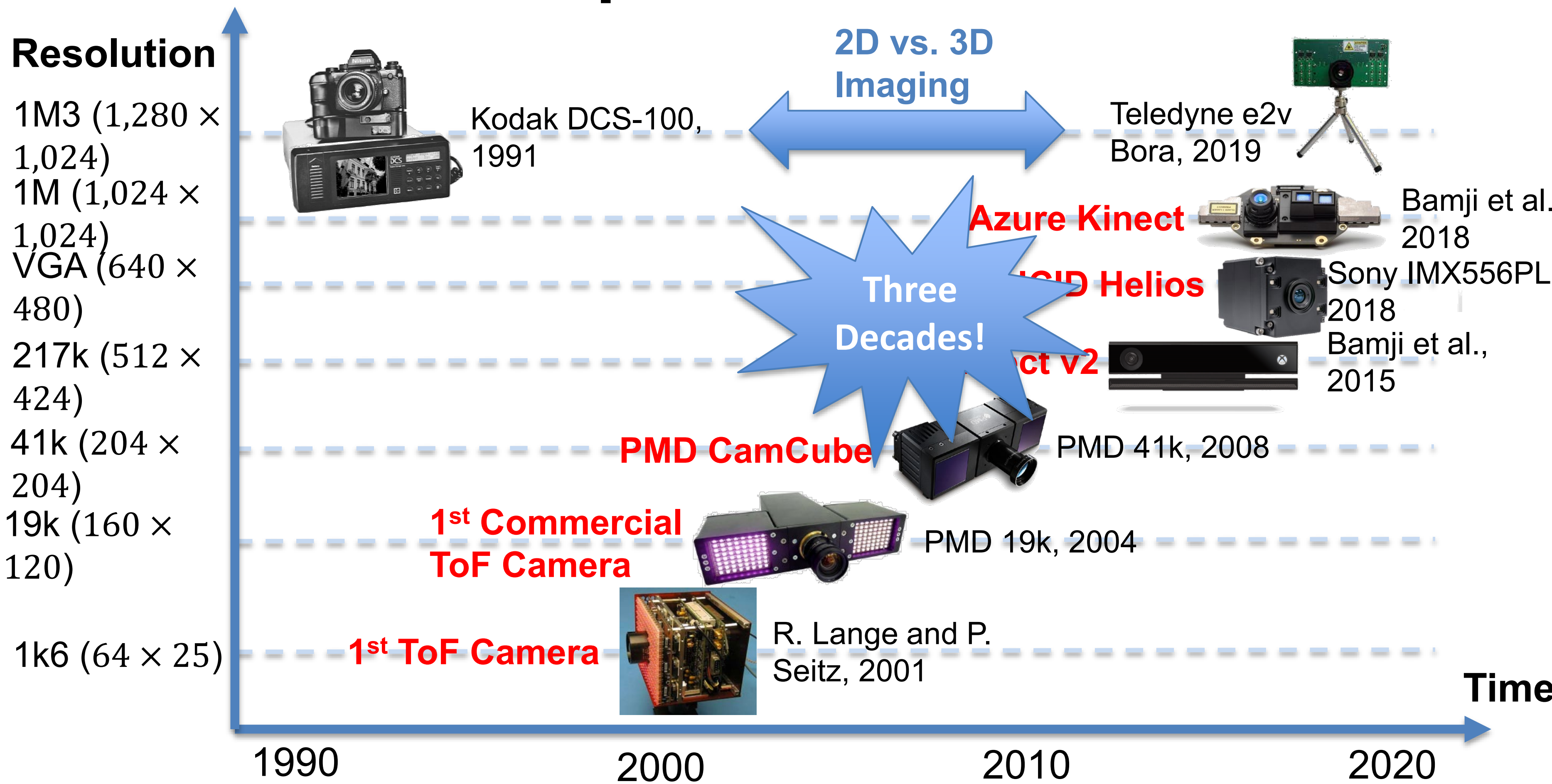
$$d = cT\underbrace{\frac{Q_B}{Q_A + Q_B}}_{\Delta t}$$

2. Why ToF Super-resolution?

The resolution problem

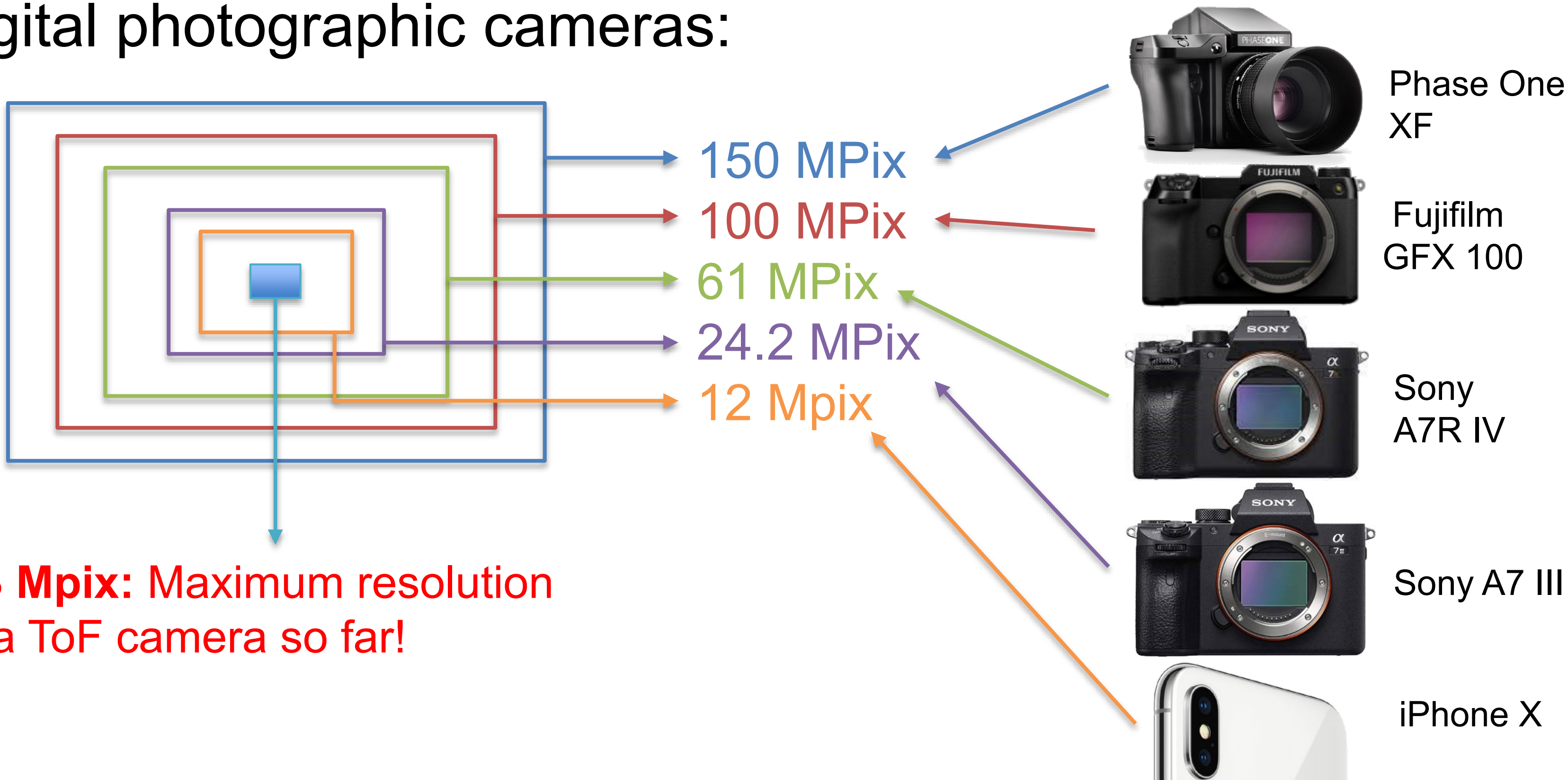
- ToF cameras do not match the resolution of conventional cameras. Reasons:
  - Active operation at a single wavelength massively restricts the collected optical power.
  - Endowing the pixels with demodulation capabilities reduces the overall pixel efficiency due to additional losses.
  - As frequency increases, demodulation contrast decreases.
  - For these reasons, larger photosensitive areas are required.
  - ToF pixels require in-pixel circuitry. This increases the pixel size and reduces the fill factor.
  - **Result:** arrays of lower resolution for the same chip area.

The resolution problem



The resolution problem

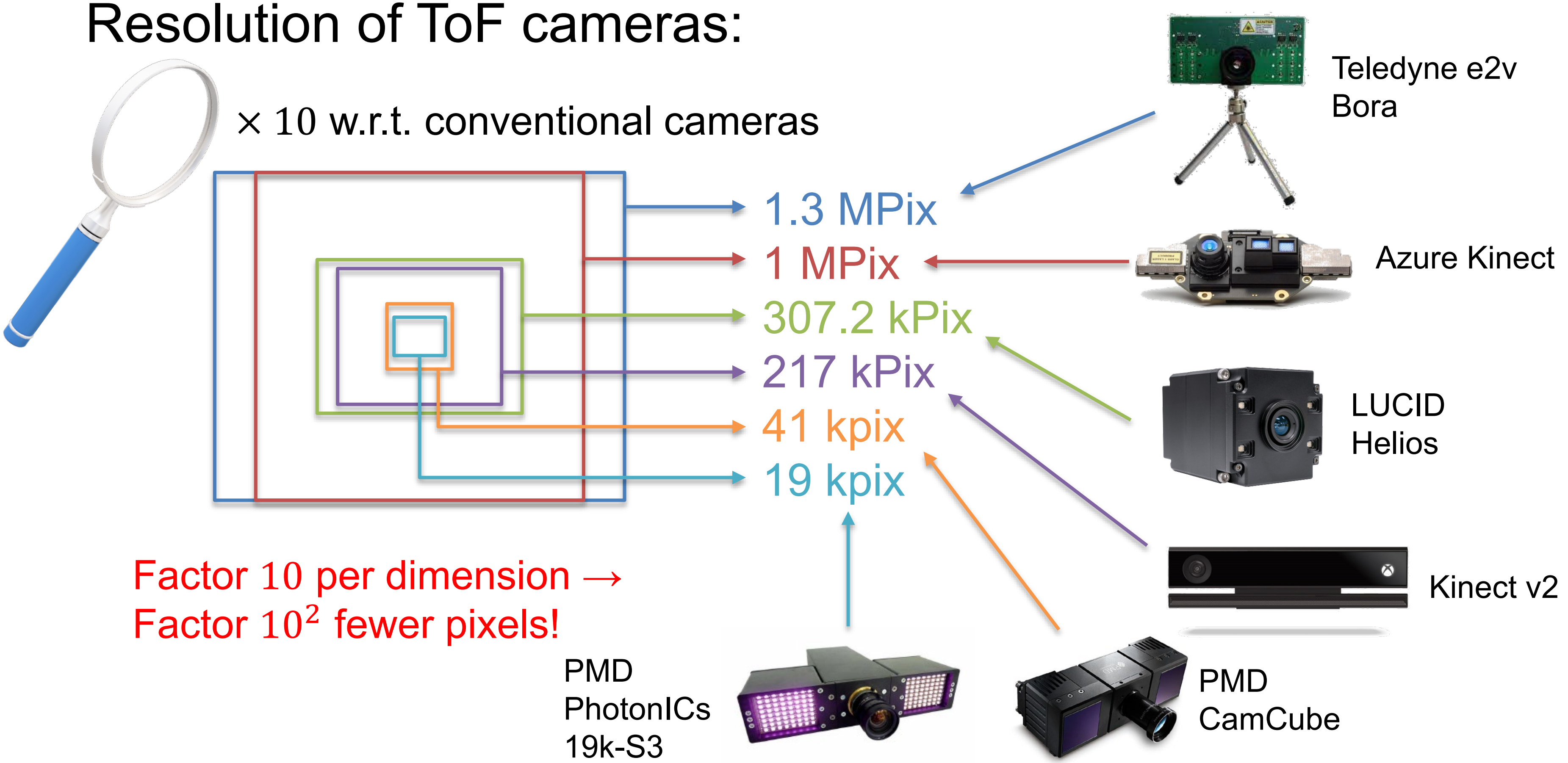
Resolution of conventional 3:2 (full frame/APS-C) digital photographic cameras:



1.3 Mpix: Maximum resolution of a ToF camera so far!

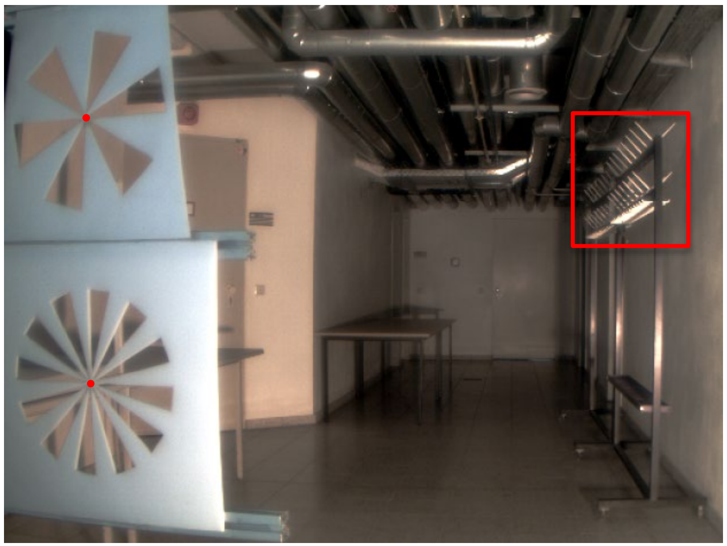
The resolution problem

Resolution of ToF cameras:



Factor 10 per dimension $\rightarrow$
Factor 10^2 fewer pixels!

The resolution problem



RGB image from a ZESS MultiCam (Aptina, 3MPix)

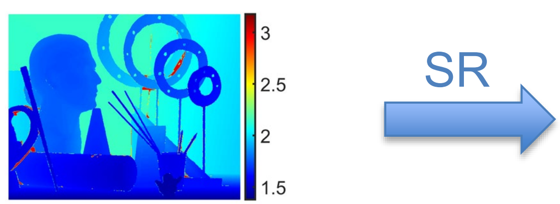
Depth image from a ZESS MultiCam (PMD PhotonICs 19k-S3)

- A ZESS MultiCam was used to simultaneously deliver registered RGB and depth images of the same scene, but…
- The **resolution gap** makes data fusion challenging.

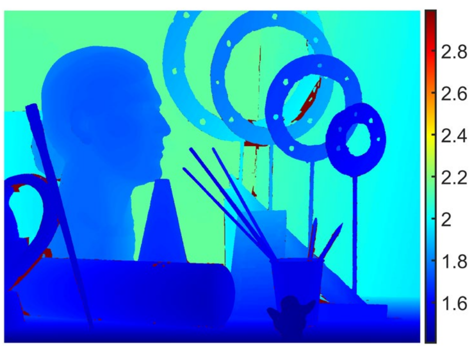
3. Methods for ToF Super-resolution

How to close the resolution gap?

- Existing literature on ToF super-resolution (SR) can be classified in three groups:
  a) Resolution transfer approaches
  b) Single-frame SR approaches
  c) Multiple-frame SR approaches



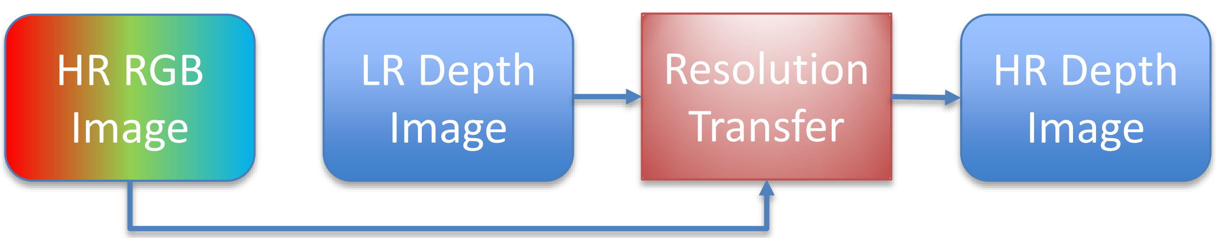
Low Resolution (LR) Depth Image



High Resolution (HR) Depth Image

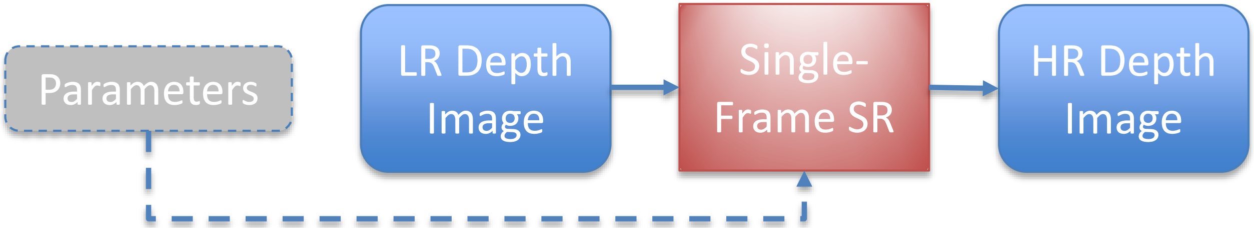
How to close the resolution gap?

- Resolution Transfer:
  – Another modality of higher-resolution is required, typically an RGB image
  – Perfect registration is needed
  – Existing methods are based on:
    - Bilateral filters
    - Markov Random Fields (MRF)
    - Neural networks



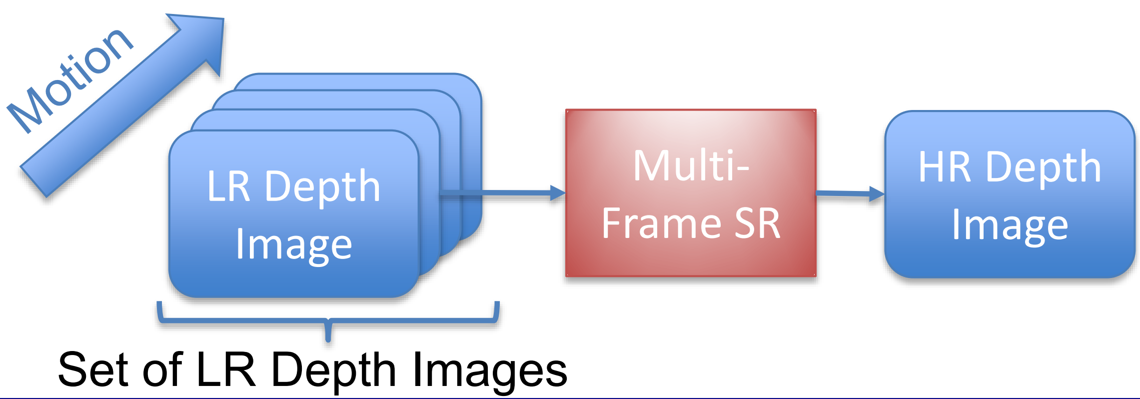
How to close the resolution gap?

- Single-frame Super-resolution:
  - It does not require another modality
  - Existing methods are based on:
    - Interpolation (bilinear, biquadratic, bicubic, etc.)
    - Deconvolution. Challenges:
      - Typically blind
      - Depth- (3D) dependent blur kernel $\rightarrow$ No (3D) shift invariance
      - Ill-posed problem $\rightarrow$ Regularizers needed
    - Neural networks

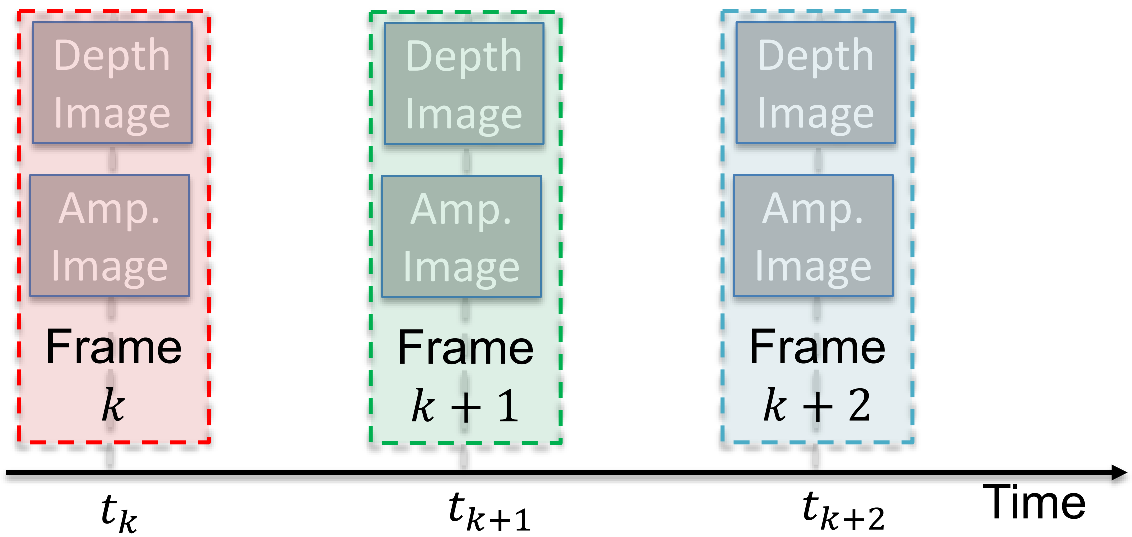


How to close the resolution gap?

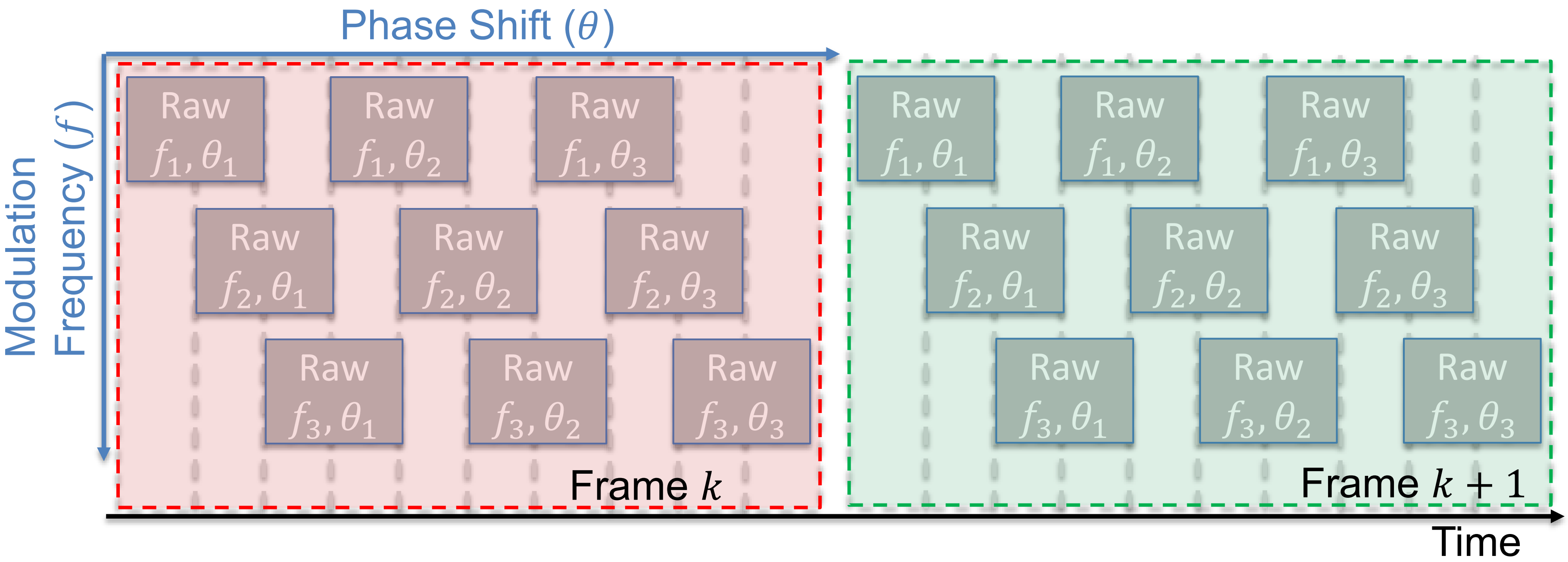
- Multiple-frame Super-resolution:
  - Leverages motion between camera and scene
  - LR depth images acquired from slightly different viewpoints are used to generate a HR depth image
  - **Main weakness:** motion is supposed to occur between frames but not within each frame!

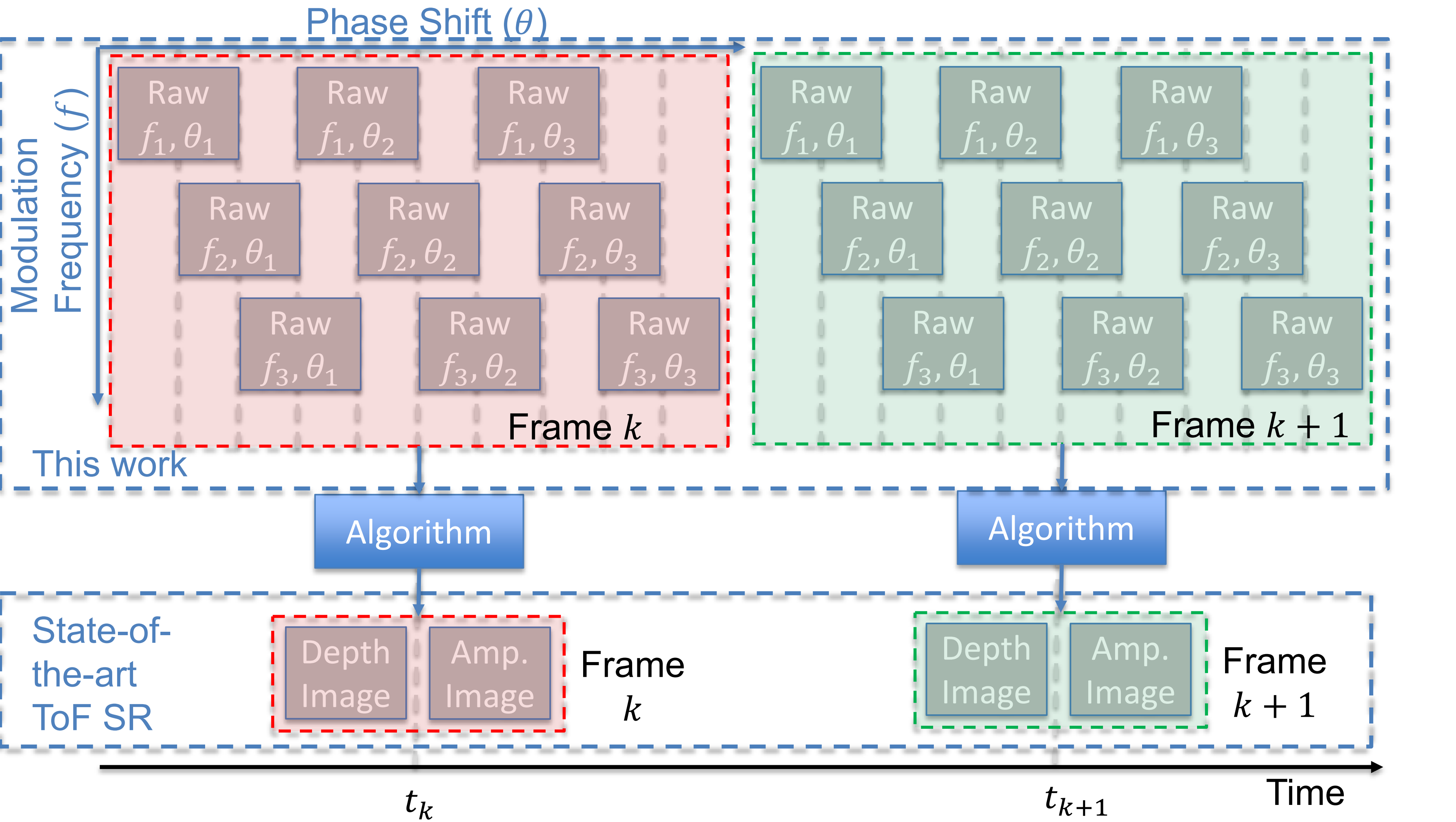


- Intra-frame motion has so-far been ignored in ToF SR.
- In practice, if motion exists, it will be both inter-frame and intra frame. → Existing multi-frame SR will then fail.
- Existing approaches rely on the underlying hypothesis that depth images are “acquired” within a negligible time window:



- In reality, amplitude and depth images are generated from sets or raw images, acquired sequentially.
- For a CW-ToF camera acquiring three phases at three different frequencies, such as the Kinect v2:





Phase Shift (θ)
Modulation Frequency (f)
Raw f_1, θ_1
Raw f_1, θ_2
Raw f_1, θ_3
Raw f_2, θ_1
Raw f_2, θ_2
Raw f_2, θ_3
Raw f_3, θ_1
Raw f_3, θ_2
Raw f_3, θ_3
Frame k
Frame $k+1$
This work
Algorithm
Algorithm
State-of-the-art ToF SR
Depth Image
Amp. Image
Frame k
Frame $k+1$
t_k
t_{k+1}
Time

4. Intra-frame ToF Super-resolution from Raw Data

- We adopt a unifying perspective and aim to attain SR from raw images.
- Considering raw images from successive frames, our framework naturally contemplates both inter-frame and intra-frame SR.
- In this work, we focus on CW-ToF, but the pipeline is valid for pulsed ToF too.
- In CW-ToF each pixel acquires measurements of the shape:

$$m[i,j] = A\left(1 + \cos\left(2\pi f_i (t - t_0) + \theta_j\right)\right)$$

- Most ToF (e.g., PMD) pixels implement two channels, acquiring measurements with $180°$ phase shift.
- In this work we use PMD technology and denote the two pixel channels by A and B.
- In this case, the sum of both channels yields an intensity measurement, ideally independent from f and θ:

$$I[i,j] = m_{\mathrm{A}}[i,j] + m_{\mathrm{B}}[i,j] = I, \; \forall i,j$$

- Provided that I can be computed per raw image acquisition, these images can be used to estimate the **motion** between consecutive raw images.

Direct Sensing Model for SR:

- Image formation model as composition of:
  1. **Motion**, modeled by the 2D motion operator $\mathcal{M}$
  2. **Blur**, modeled by the convolution kernel $B(x, y)$
  3. **Downsampling**, modeled by the operator $\mathcal{D}$

$$m[i, j, k, \boxed{u, v}] = \mathcal{D}\left(B(x, y) *^2 \boxed{\mathcal{M}_{\boxed{i,j,k}}}(m[i, j, k, \boxed{x, y}])\right)$$

LR 2D Spatial Domain (discrete)

Common for all pixel channels!

i, j: *intra*-frame motion
k: *inter*-frame motion

HR 2D Spatial Domain (discrete/ continuous)

LR to HR - Inverting the Direct Model:

- Two-step fast and robust method [1] :
  1. Non-iterative **data fusion**
  2. Iterative **deblurring**
- Let $\underline{Z}[i,j] = \boldsymbol{B}\underline{m}^{\mathrm{HR}}[i,j]$, be the blurred version of the HR raw image we aim to estimate in step 1. Then, we seek:

$$\hat{\underline{Z}}[i,j] = argmin_{\underline{Z}} \sum_{k=0}^{K} \left\| \boldsymbol{D}\boldsymbol{M}_{i,j,k}\underline{Z} - \underline{m}[i,j,k] \right\|_p^p$$

Where $\boldsymbol{B}$, $\boldsymbol{D}$, and $\boldsymbol{M}_{i,j,k}$ are matrix equivalents of discrete operators.

Closed-form solutions:

– $p = 2$: **mean** value of *registered* frames

– $\boxed{p = 1}$: **median** value of *registered* frames

[1] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, “Fast and robust multiframe super resolution,” in: *IEEE Trans. Image Proc.*, vol. 13, no. 10, pp. 1327–1344, 2004.

LR to HR - Inverting the Direct Model:

- How to attain intra-/inter-frame **registration**? → Use the intensity images $I[i,j]$
- To obtain 2D displacements, retrieve first phase shifts in 2D-frequency domain [2]:

$$2\pi[f_{\mathrm{H}}\ f_{\mathrm{V}}]\begin{bmatrix}\Delta x_{i,j,k}\\ \Delta y_{i,j,k}\end{bmatrix} = \arg\left(\frac{\mathcal{F}_{f_{\mathrm{H}},f_{\mathrm{V}}}\boldsymbol{I}[i,j,k]}{\mathcal{F}_{f_{\mathrm{H}},f_{\mathrm{V}}}\boldsymbol{I}[i_{\mathrm{r}},j_{\mathrm{r}},k_{\mathrm{r}}]}\right)$$

2D Displacement — Phase Shift — Reference intensity image

- For a set of $f_{\mathrm{H}}\ f_{\mathrm{V}} \in \Omega^2$ (low-pass region), we obtain a set of equations. → Retrieve $\Delta x_{i,j,k},\ \Delta y_{i,j,k}$ via least squares.

[2] P. Vandewalle, S. Süsstrunk, and M. Vetterli, “A frequency domain approach to registration of aliased images with application to superresolution,” in: *EURASIP Journal on Advances in Signal Processing*, vol. 2006, no. 1, pp. 1–14, 2006.

LR to HR - Inverting the Direct Model:

- Two-step fast and robust method [1] :
  1. Non-iterative **data fusion**
  2. Iterative **deblurring**
- For the iterative deblurring we use the collaborative filtering extension of BM3D [3]. Key points:
  - **Patch matching** allows obtaining 3D data by grouping similar 2D patches
  - **Sparsity** is pursued in a 3D transform domain
  - Plus: exact computation of the noise variance in transform domain
  - **Regularized inversion:** the attained noise reduction compensates the inherent noise amplification of a low-pass deconvolution (deblurring)

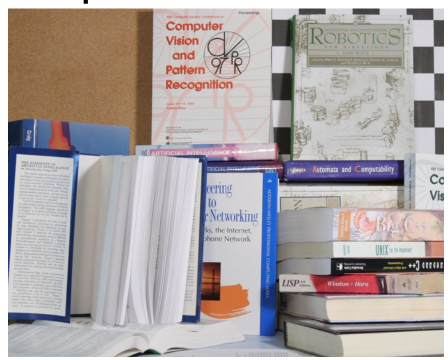
[3] Y. Mäkinen, L. Azzari, and A. Foi, "Collaborative filtering of correlated noise: Exact transform-domain variance for improved shrinkage and patch matching," in: *IEEE Transactions on Image Processing*, vol. 29, pp. 8339–8354, 2020.

5. Experimental Results

Synthetic Experiments:

- Datasets: **Middlebury** stereo datasets 2003 and 2005:
  - RGB + disparity of 8 complex scenes. **RGB** samples from 2005:

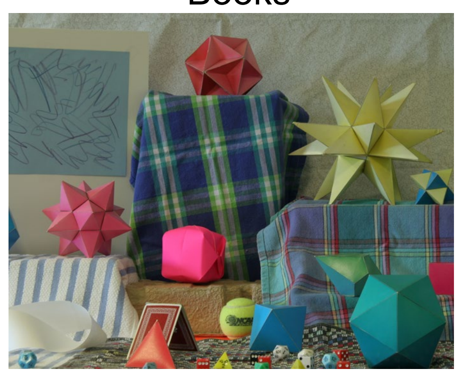
Art



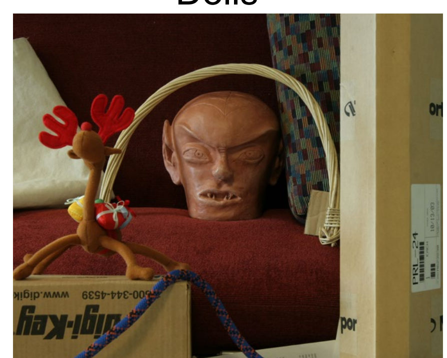
Books

Dolls

Laundry



Moebius



Reindeer

Synthetic Experiments:

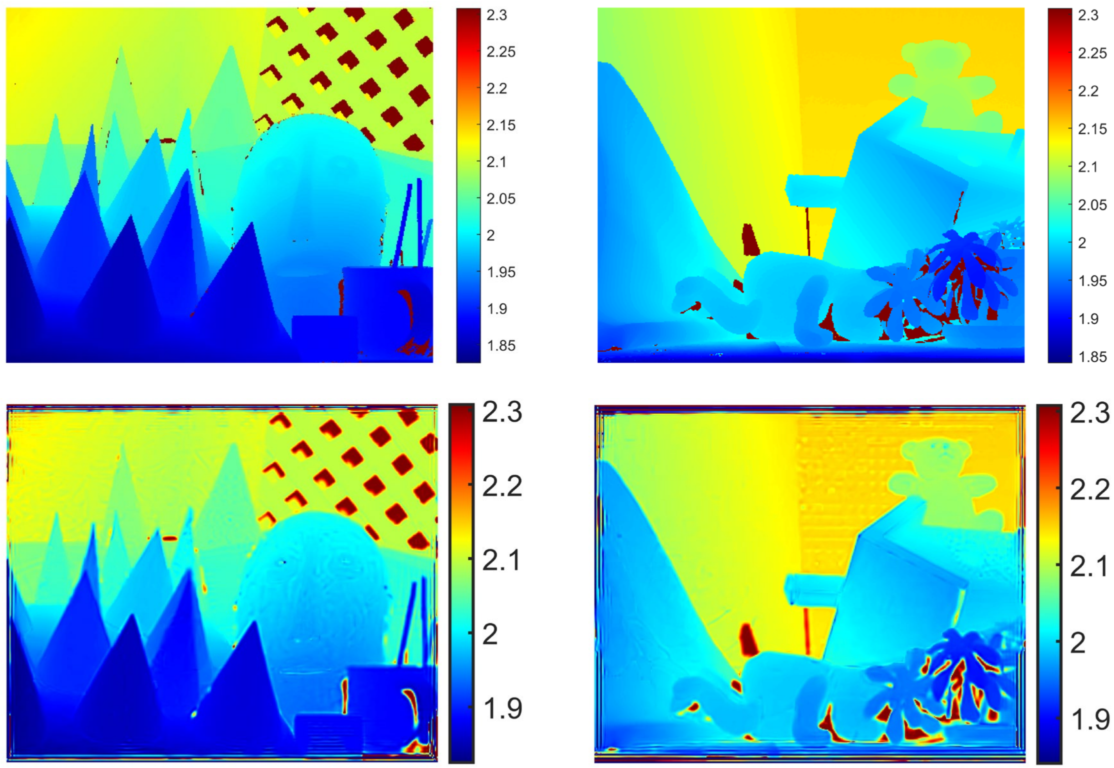
- Datasets: **Middlebury** stereo datasets 2003 and 2005:
  - RGB + disparity of 8 complex scenes.
- For each scene, 15 frames of HR synthetic ToF raw data (2 pixel channels, 4 phases, 1 frequency) are generated.
- Raw images of both pixel channels are randomly 2D-shifted, up to ± 5 pixels in HR domain.
- 10 independent experiments per scene.
- The shifted HR raw data is blurred and downsampled to generate the LR raw data.
- Apply the proposed SR pipeline to LR raw data (SR factor 2).
- HR depth images are obtained via the *four phases algorithm.*

Synthetic Experiments. Results:

- **Middlebury** stereo dataset 2003:

HR Ground Truth

Depth from SR Raw Data



Cones

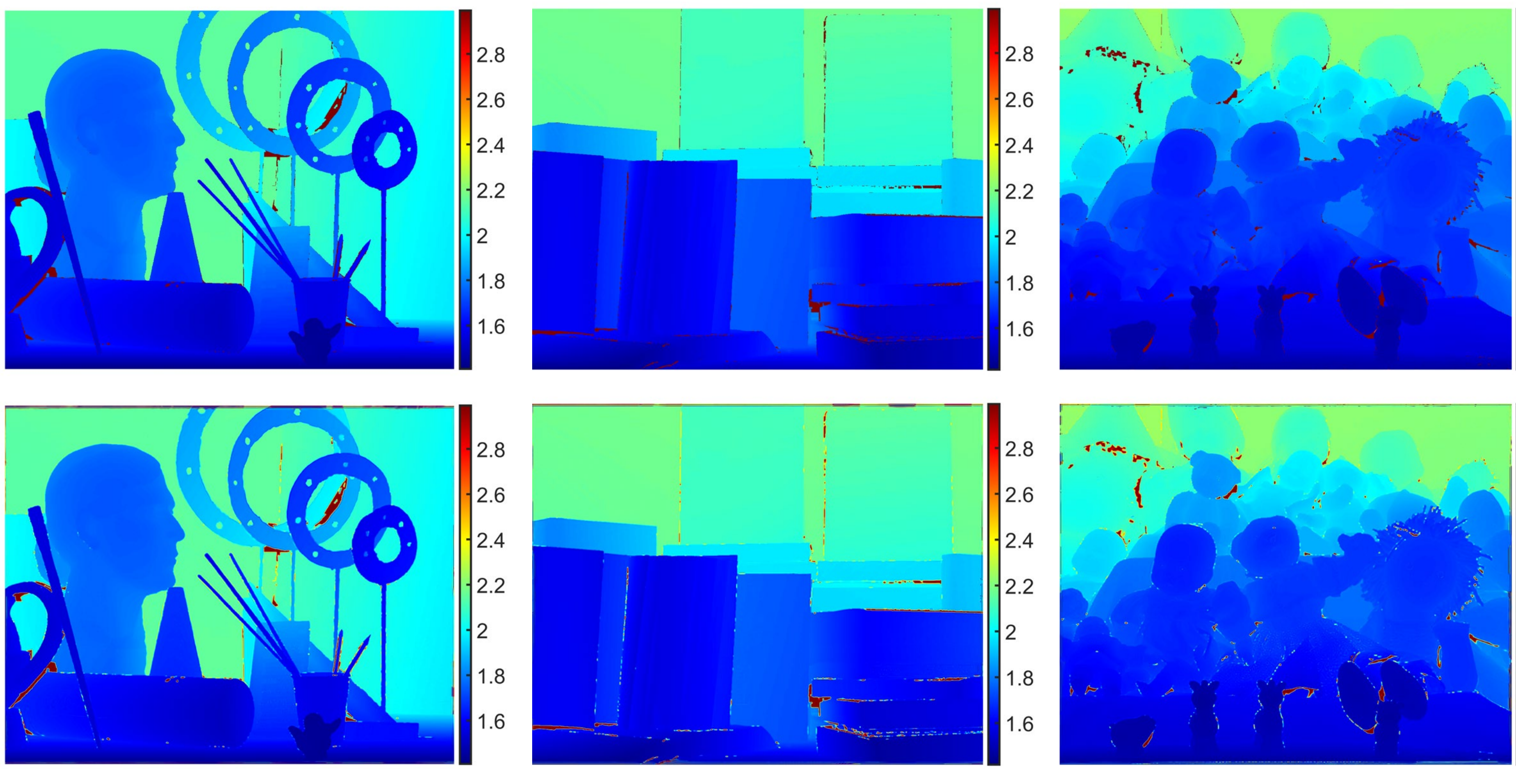
Teddy

Synthetic Experiments. Results:

- **Middlebury** stereo dataset 2005:

HR Ground Truth

Depth from SR Raw Data



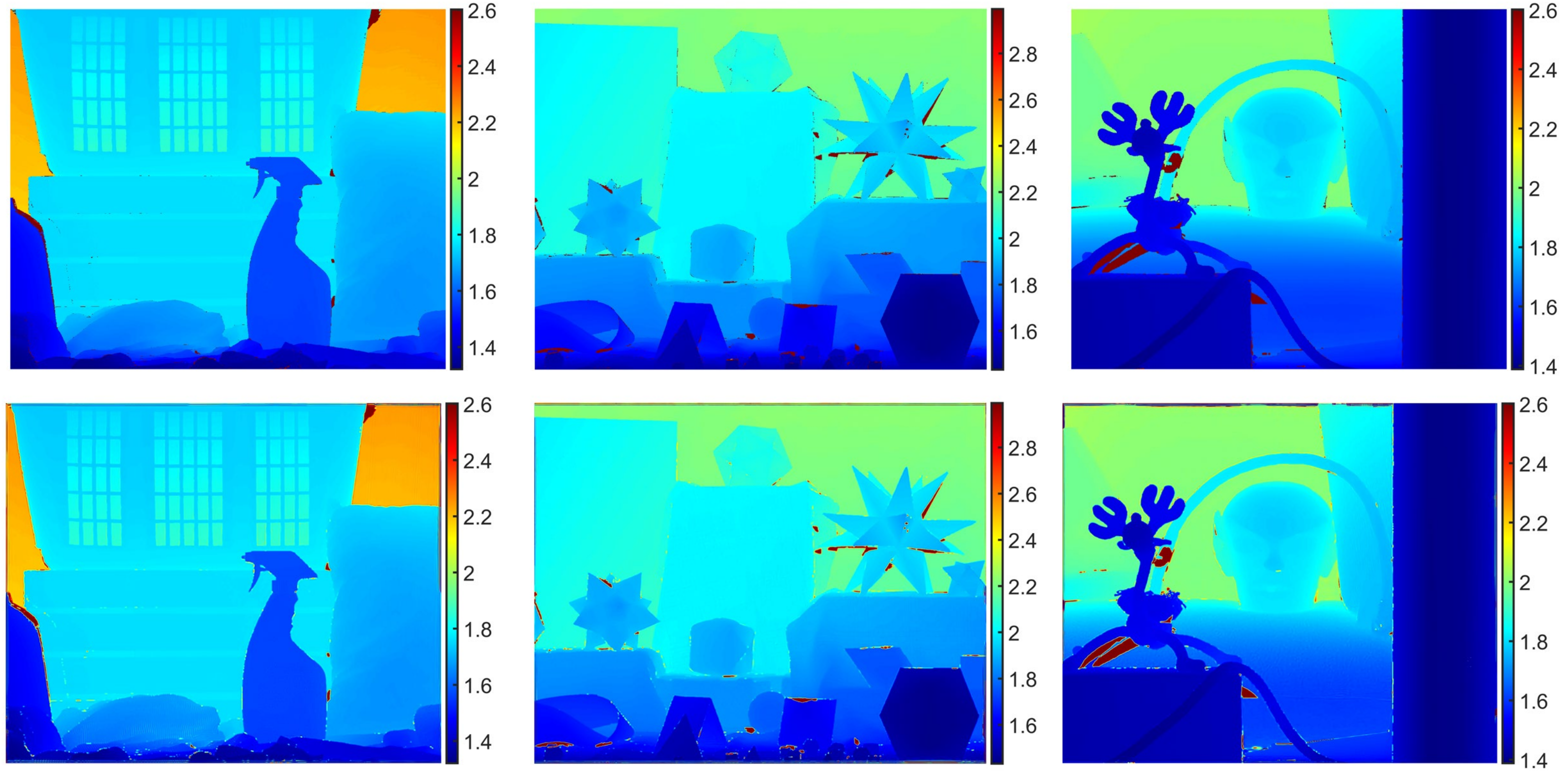
Art | Books | Dolls

Synthetic Experiments. Results:

- **Middlebury** stereo dataset 2005:

HR Ground Truth

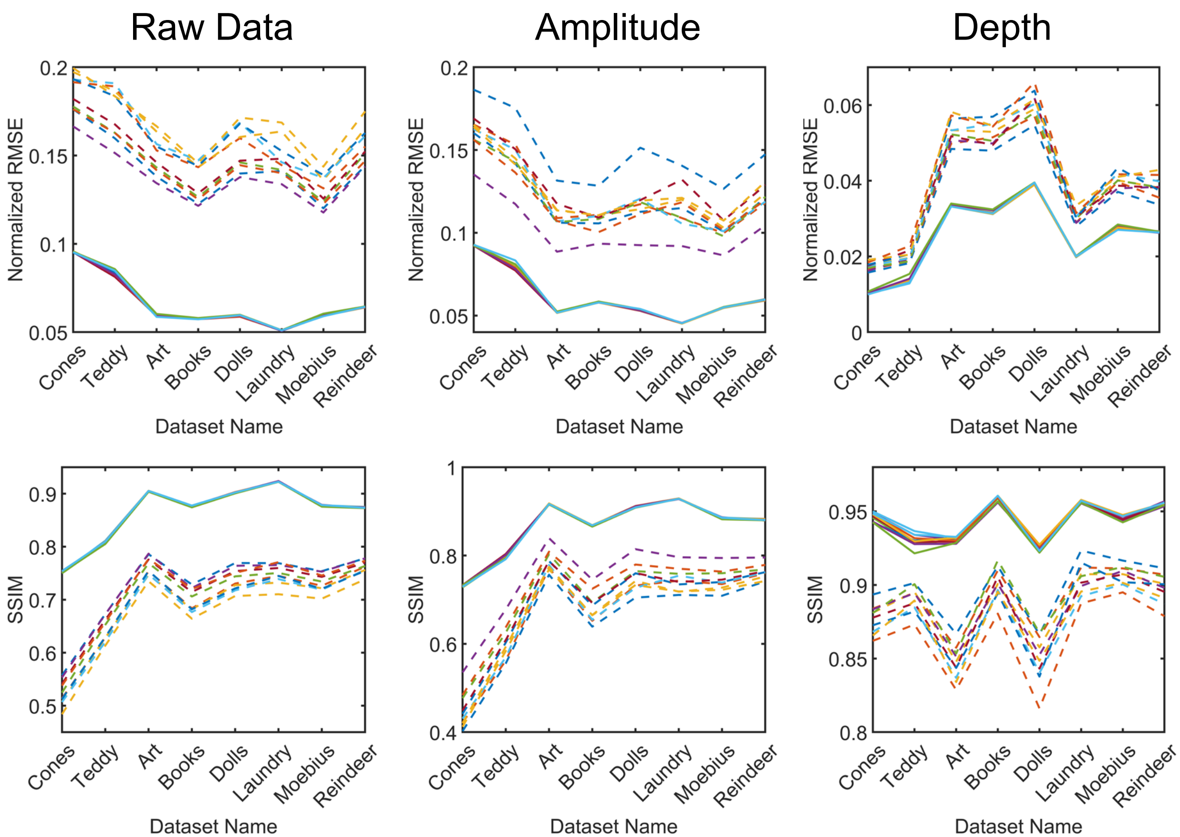
Depth from SR Raw Data



Laundry Moebius Reindeer

Synthetic Experiments. Results:

- **Middlebury** stereo datasets. RMSE and SSIM plots:

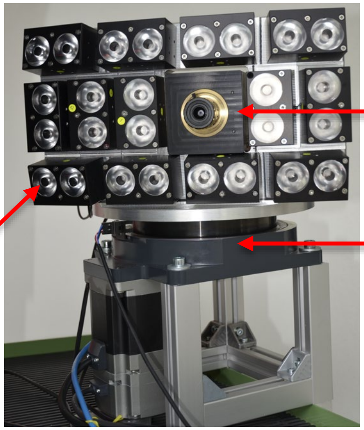


- 8 scenes
- 10 experiments
- <u>Solid:</u> SR result
- <u>Dashed:</u> bicubic interpolation

Real Experiments:

- **Hardware: ZESS MultiCam** with medium-range illumination system mounted on a rotary table
- Accurate angular control allows for custom (horizontal) displacements with subpixel accuracy
- Test scene: hall of ZESS building (≥16.5m range)
- Two horizontal inter-frame displacements considered:
  a) 1.34 pixels
  b) 6.43×10^{-2} pixels
- 15 consecutive raw data frames

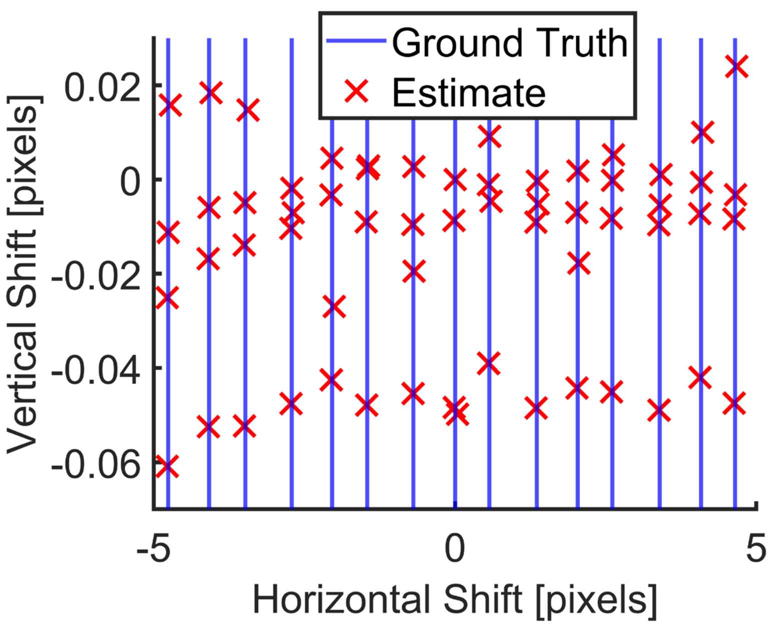
ZESS Hall



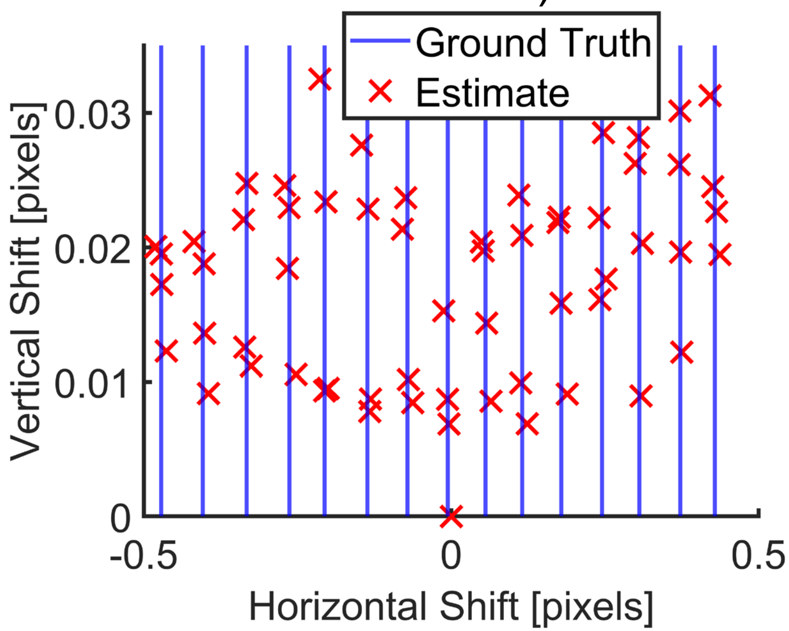
Real Experiments. Registration Results:

- The proposed raw image registration procedure attains high subpixel accuracy, e.g., in the order of $\mathbf{10^{-3}}$ **pixels** in case b)

Case a):



Case b):



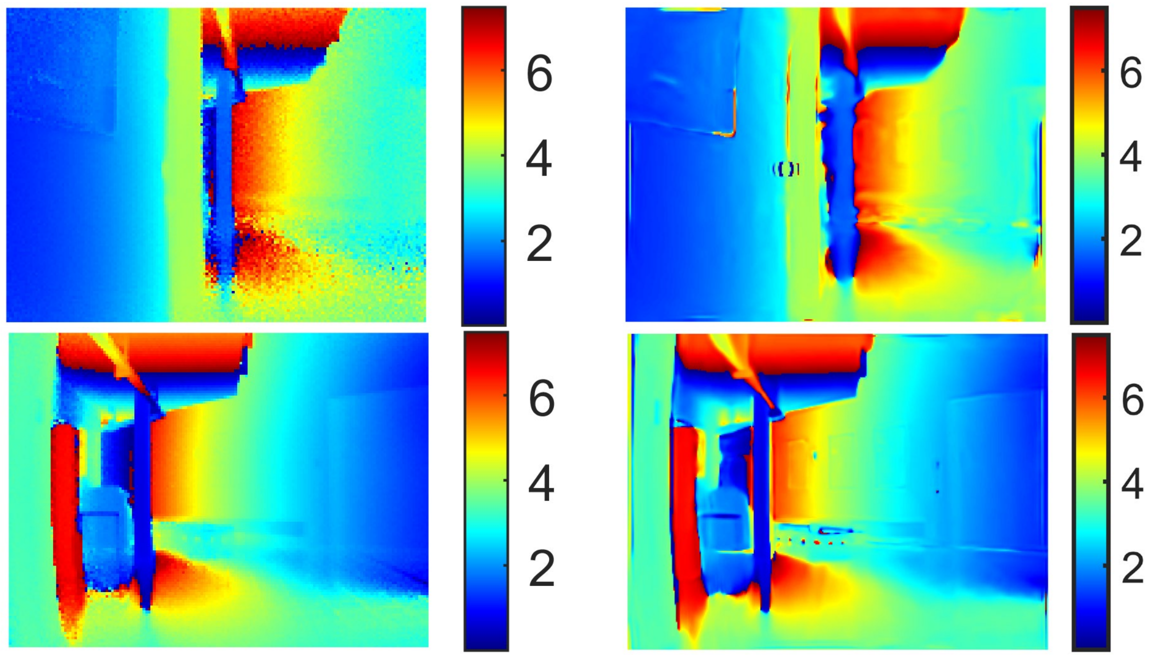
Real Experiments. Depth SR Results:

- The acquired raw data is used as input for our SR pipeline. The SR raw data is then used to obtain a depth image.

Depth from LR Raw Data | Depth from SR Raw Data

Case a):

Case b):



6. Conclusions

Conclusions:

- ToF cameras can retrieve 3D, but its resolution is an order of magnitude lower than conventional 2D cameras.
- Existing multi-frame SR methods ignore intra-frame motion and operate directly on depth images.
- We have presented a SR framework that works on ToF raw data and accounts for both inter- and intra-frame motion.
- Based on two separable tasks:
  - Raw data fusion
  - Deblurring
- Experiments on synthetic and real ToF data from challenging scenes witnessed good performance of the approach.

Thank you for
your Attention!

Do not hesitate forwarding your questions to:
heredia@zess.uni-siegen.de