Image-Guided ToF Depth Upsampling

I. Eichhardt, D. Chetverikov, and Z. Jankó

MTA SZTAKI, Budapest, Hungary

SZTE, 5-May-2015

Outline

1. Introduction
2. Time-of-Flight cameras
3. Image-guided depth upsampling
    - Problems of image-guided depth upsampling
    - Local methods
    - Global and other methods
4. Conclusion

Outline

Image-based 3D reconstruction

- Early years of computer vision (CV)
  - humans perceive dynamic 3D scenes based on vision
  - → visual info sufficient for computer to reconstruct scene
- Today's view
  - humans do not need to build precise 3D models
  - → to be able to act in dynamic world
  - many CV applications need precise reconstruction
  - → fuse data by different sensors
  - → combine different methods

Techniques to acquire depth

- Stereo vision
  - calibrated stereo rig
  - multiview
- Photometric stereo
  - multiple light sources
- Shape-from-X
  - motion, texture, shading, focus
  - less precise, less robust
- Structured light
  - Kinect
- Direct depth measurement
  - LIDAR devices: 3D point clouds
  - Time-of-Flight (ToF) cameras: depth images

Outline

Features of ToF cameras

- We only consider widely used **lock-in** ToF cameras
  - rarely used pulse-based ToF cameras ignored here
- Main features
  - small, compact
  - low weight
  - low consumption
  - no moving parts (contrary to LIDAR)
- Principle of operation
  - emit infrared light
  - measure time-of-flight to observed object
- ToF camera delivers
  - depth images at video frame rates
  - registered reflectance images of same size
  - reliability values of depth measurements

Main disadvantages of ToF cameras

- Low resolution
  - highest current resolution: QVGA (320×240)
  - target of future development: VGA (640×480)

  $\rightarrow$ limited by chip size
- Significant acquisition noise
  - low quality depth images
  - quality gradually improving

  $\rightarrow$ limited by small active illumination energy
- Problems with outdoor use
  - bright lighting can increase ambient light noise in ToF

  $\rightarrow$ if ambient light contains same wavelength as camera light

Applications of ToF cameras

- Most exploited features of ToF cameras
  - ability to operate without moving parts
  - providing depth maps at high frame rates

  $\rightarrow$ greatly simplifies foreground-background separation
- Mobile robot vision
  - navigation
  - 3D pose estimation and mapping
- 3D reconstruction of objects and environments
- Computer graphics and 3D television (3DTV)
- Recognition and tracking of people and parts of body
  - face, hands

Comparison of four depth acqusition techniques

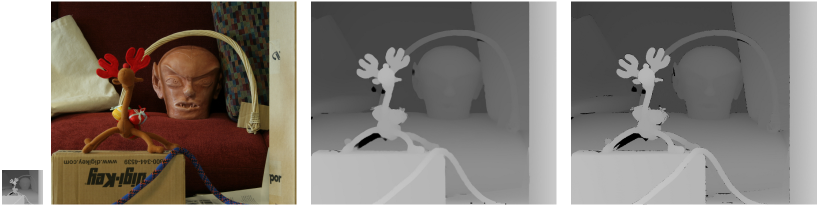
feature	stereo	phot. stereo	struct. light	ToF camera
correspond.	yes	no	yes	no
extrinsic calib.	yes	yes	yes	no
active illum.	no	yes	yes	yes
weak texture	weak	good	good	good
strong texture	good	medium	medium	medium
low light	weak	good	good	good
bright light	good	weak	med./weak	medium
outdoor	yes	no	no	yes?
dynamic	yes	no	yes	yes
resolution	cam.dep.	cam.dep.	cam.dep.	low
accuracy	mm to cm	mm	μm to cm	mm to cm

Outline

Types of depth image upsampling

- Goal: increase resolution, improve image quality
  - in particular, at depth edges
  - $\rightarrow$ erroneous or lacking measurements
- Upsampling with stereo: **ToF–stereo** fusion
  - recent survey available
  - ToF and passive stereo complement each other
  - $\rightarrow$ fuse two sources of depth data
  - problems: matching, textureless surfaces, occlusions
- Fusing **multiple ToF depth measurements**
  - does not need sensor of other type
  - most methods applicable to static scenes only
- Upsampling based on **single image**
  - $\rightarrow$ this talk

Example of upsampling: Middlebury data



input data | upsampled depth | ground-truth depth

- Middlebury stereo dataset: depth and colour images
  - original high-resolution depth: ground truth (structured light)
  - downsampled to low-resolution to get input depth
- Good result (Ferstl 2013) for non-realistic high-quality data
  - upsampled depth smooth and similar to ground truth
  - small parts of depth data lost (dark regions)

Example of upsampling: real data captured in studio

input data

resized depth

upsampled depth

- Real data captured in studio[1]
  - part of data lost due to very low resolution
  - some shapes, e.g., heads and hands, distorted
  - contours of depth and colour images do not always coincide
  - erroneous and lacking measurements, esp. at depth edges
- Reasonably good result (Jankó 2014) for real input
  - except for a few problematic areas

[1] Data courtesy of Zinemath Zrt.

Effect of imprecise calibration on upsampling

2 pixels | 5 pixels | 10 pixels

- Illustrated by artificially introduced discrepancy
  - → shift between input depth and colour images
  - shift magnitude 2, 5, 10 pixels
  - upsampled depth borders become blurred and coarse

Texture transfer problem

colour image | ups.depth (Ferstl 2013) | ground truth

- Basic assumptions of most upsampling methods:
  1. intensity edges coincide with depth edges
  2. smooth intensity $\rightarrow$ smooth surface, depth
- First assumption violated in areas of strong texture
  - contrast image textures ‘imprint’ onto upsampled depth
  - $\rightarrow$ methods have to relax assumption
- Principal question: How to combine depth and colour?

Notations

$\mathtt{D}$	Input (depth) image
$\hat{\mathtt{D}}$	Filtered / Upsampled image
$\nabla\mathtt{D}$	Gradient image
$\tilde{\mathtt{I}}$	Guide / Reference image
$p, q, \ldots$	2D pixel coordinates (vectors)
$p_{\downarrow}, q_{\downarrow}, \ldots$	low-resolution coordinates (possibly, fractional)
$\Omega(p)$	Window around pixel p
$f, g, h, \ldots$	Gaussian kernel functions
k_p	Normalisation factor: sum of weights in window around pixel p (possibly, location-dependent)

Filter definitions

- **Convolution** with location-dependent weights

$$\hat{\mathtt{D}}_p = \frac{1}{k_p} \sum_{q \in \Omega(p)} W(p, q)\, \mathtt{D}_q, \quad \text{where} \quad k_p = \sum_{q \in \Omega(p)} W(p, q)$$

- **Bilateral** filter: kernels with photometric & spatial distances

$$W_B(p, q) = f(\|\mathtt{D}_p - \mathtt{D}_q\|)\, g(\|p - q\|)$$

- **Joint bilateral** filter[2]: guidance image $\tilde{\mathtt{I}}$ of same size as $\mathtt{D}$

$$W_{JB}(p, q) = f(\|\tilde{\mathtt{I}}_p - \tilde{\mathtt{I}}_q\|)\, g(\|p - q\|)$$

[2] Also called **cross bilateral**

Joint bilateral upsampling

- Depth image $\mathtt{D}$ smaller than optical guidance image $\tilde{\mathtt{I}}$
  - $\rightarrow$ filters operate with low-resolution pixel coordinates $q_\downarrow$
  - $\rightarrow$ depth interpolation for fractional coordinates
- **Joint Bilateral Upsampling** filter:

$$\hat{\mathtt{D}}_p = \frac{1}{k_p} \sum_{q_\downarrow \in \Omega(p_\downarrow)} W_{JBU}(p, q)\, \mathtt{D}_{q_\downarrow}$$

$$W_{JBU}(p, q) = f\left(\|\tilde{\mathtt{I}}_p - \tilde{\mathtt{I}}_q\|\right) g\left(\|p_\downarrow - q_\downarrow\|\right)$$

Joint bilateral median upsampling

- **Weighted median** filter:

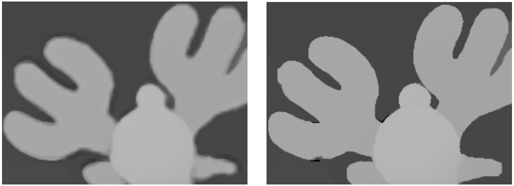
$$\hat{\mathrm{D}}_p = \arg\min_b \sum_{q \in \Omega(p)} W(p, q) \left| b - \mathrm{D}_q \right|$$

  - median minimises total weighted photometric distance to other pixels of window
  - definition equivalent to traditional one (via sorting)

- **Joint Bilateral Median Upsampling** filter:

$$\hat{\mathrm{D}}_p = \arg\min_b \sum_{q_\downarrow \in \Omega(p_\downarrow)} W_{JBU}(p, q) \left| b - \mathrm{D}_{q_\downarrow} \right|$$

Comparison of JB and JBM upsamplings



Joint Bilateral Joint Bilateral Median

- Weighted average vs weighted median, bilateral weights
  - → gradual blending vs sudden transitions
  - → finer variation in depth vs better ‘segmentation’
- JB upsampling follows colour variations
  - → likely to result in depth interpolation
- JBM is more robust to colour variations and outliers
  - → less depth interpolation
  - → less texture transfer (copying)

Pixel Weighted Average Strategy

$$\hat{\mathrm{D}}_p = \frac{1}{k_p} \sum_{q \in \Omega(p)} h\left(\left|\nabla \mathrm{D}_{q_\downarrow}\right|\right) W_{JBU}(p, q)\, \mathrm{D}_{q_\downarrow}$$

- $h\left(\left|\nabla \mathrm{D}_{q_\downarrow}\right|\right)$: **credibility map**
  - h: Gaussian kernel
  - $\left|\nabla \mathrm{D}_{q_\downarrow}\right|$: depth gradient magnitude
- Prefers locations of small depth changes
  - $\rightarrow$ average over smooth surfaces
  - $\rightarrow$ do not average across depth edges

Noise-aware filter for depth upsampling

$$\hat{\mathtt{D}}_p = \frac{1}{k_p} \sum_{q_\downarrow \in \Omega(p_\downarrow)} g\left(\|p_\downarrow - q_\downarrow\|\right) \cdot \left[\alpha_p f_1\left(\|\tilde{\mathtt{I}}_p - \tilde{\mathtt{I}}_q\|\right) + (1 - \alpha_p)\, f_2\left(\left\|\mathtt{D}_{p_\downarrow} - \mathtt{D}_{q_\downarrow}\right\|\right)\right] \mathtt{D}_{q_\downarrow}$$

- Composite Joint Bilateral filter
- Via α_p, locally adapts to:
  - noise level of depth function
  - smoothness of depth function
- Depending on local context, blends:
  - standard JBU ($\alpha_p = 1$)
  - edge-preserving smoothing depth filter independent from colour data ($\alpha_p = 0$)

Non-Local Means: generalised bilateral

- Photometric term in bilateral similarity kernel
  - BF: point-wise intensity difference
  - NLM: patch-wise intensity difference
- Geometric term in bilateral similarity kernel
  - BL: distance between points
  - NFM: distance between patches
- NLM allows for large (in principle, infinite) distances
  - $\rightarrow$ strong contribution from distant patches
  - $\rightarrow$ non-local
- In practice:
  - search for patches limited to reasonable neighbourhood
  - $\rightarrow$ more or less local
  - photometric term gives higher importance to patch centers
  - $\rightarrow$ Gaussian weights to distant patch pixels

Markov Random Field based upsampling

- Diebel 2005: Two-layer MRF
  - quadratic difference between measured & estimated depths
  - depth smoothing prior
  - weighting factors relating image edges to depth edges
  - $\rightarrow$ least square optimisation problem
  - $\rightarrow$ solved by conjugate gradient algorithm
- Lu 2011
  - linear cost term (truncated absolute difference)
  - $\rightarrow$ more robust to outliers
  - adaptive elements in cost term
  - $\rightarrow$ solved by loopy belief propagation
- Choi 2012
  - quadratic cost terms
  - discrete and continuous optimisation
  - multiresolution framework

Optimisation based upsampling without MRF

- Cost terms often similar to those used for MRF
- Ferstl 2013
  - energy functional: standard quadratic depth data term **&**
  - regularising Total Generalized Variation (TGV) term **&**
  - anisotr. diffusion term relating image and depth gradients **&**
  - $\rightarrow$ energy minimised by primal-dual optimisation algorithm
  - MATLAB source code available at project web site
- Park 2011
  - energy functional includes NLM regularising term
  - helps preserve local structure and fine details
  - $\rightarrow$ in presence of significant noise

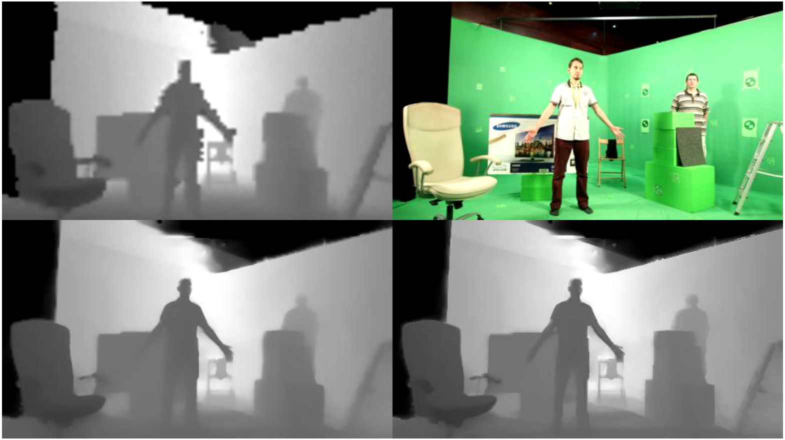
Other upsampling methods

- Tallon 2012
  - joint segmentation of depth and intensity
  - → regions of homogeneous colour and depth
  - conditional mode estimation
  - → detect and correct regions with inconsistent features
- Soh 2012
  - oversegment colour image
  - → image super-pixels
  - → depth edge refinement
  - MAP-MRF framework to further enhance depth
- Li 2012: piecewise planar scene assumed
  - Bayesian approach to depth upsampling
  - account for intrinsic camera errors
  - simulate uncertainty of depth and colour measurements
  - use RANSAC to select inliers for each plane model
  - optimise objective function to refine depth

Video-based depth upsampling

- Based on same assumptions and principles as single-image
- Use additional techniques and constraints
  - motion-compensated frame interpolation
  - extension of 2D bilateral filter in space and time
  - other forms of spatio-temporal filtering
  - temporal coherence of depth video
  - optical flow for motion estimation
  - fast GPU implementation

Video of depth upsampling in studio



Upsampling algorithms with optical flow (Eichhardt 2015)

Outline

Outlook 1

- Depth upsampling extends beyond ToF data processing
  - point cloud upsampling
  - $\rightarrow$ similar principles and techniques
- In near future, ToF cameras will undergo fast changes
  - higher resolution and quality of depth image
  - increased measurement range
  - enhanced robustness
  - $\rightarrow$ outdoor use
- Application areas of ToF cameras will extend and grow
  - much more frequent use
  - $\rightarrow$ lower prices

Outlook 2

- Sensor data fusion becomes more and more popular
  - → trend of coupling ToF with other sensors will persist
  - → growing demand for studies in depth data fusion
- Critical issue: evaluation and comparative testing
  - tests on Middlebury dataset not particularly indicative
  - → of performance in real conditions and applications
  - needed: good, rich **benchmark** of ToF data
  - → acquired in different real-world conditions