ilmedia

TECHNISCHE UNIVERSITÄT
ILMENAU

Lietz, Henrik; Franzen, M.; Eberhardt, Jörg; Sinzinger, Stefan:

Optimized deep learning algorithms for application with data from PMD cameras

Zuerst erschienen in:	DGaO-Proceedings. - Erlangen-Nürnberg : Dt. Gesellschaft für angewandte Optik. -119 (2018), art. P28, 1 S.
Erstveröffentlichung:	25.06.2018
ISSN:	1614-8436
URN:	urn:nbn:de:0287-2018-P028-5
[*Gesehen:*	30.08.2019]

TU Ilmenau | Universitätsbibliothek | ilmedia, 2019
http://www.tu-ilmenau.de/ilmedia

Optimized deep learning algorithms for application with data from PMD cameras

TECHNISCHE UNIVERSITÄT ILMENAU

H. Lietz*, M. Franzen*, J. Eberhardt* and S. Sinzinger**

* Faculty T - Technology and Management, University of Applied Sciences Ravensburg-Weingarten

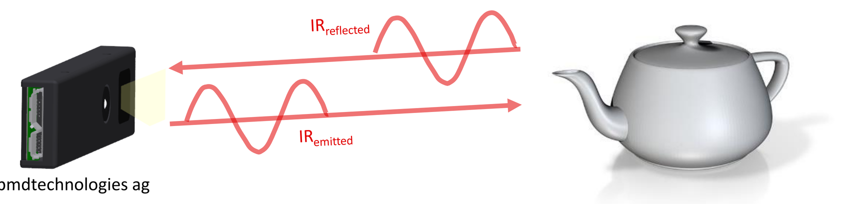
** Department of Mechanical Engineering, Group of Optical Engineering, Ilmenau University of Technology

Hochschule Ravensburg-Weingarten University of Applied Sciences

1. Motivation: super-resolution (SR) on inherently related sensor data

Time-of-Flight (ToF) Photonic Mixing Device (PMD) camera

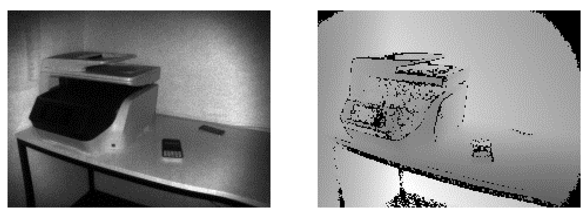
- fast and robust three-dimensional image acquisition
- PMD sensor measures the phase difference between an emitted and its reflected amplitude modulated IR signal in real time



Source: pmdtechnologies ag

Problem

- large pixel sizes limit lateral resolution
- existing depth map SR fusion approaches require a further sensor's additional high-resolution (HR) intensity image



Amplitude and distance images from PMDtec's miniaturized PMD camera PicoFlexx.

Goal

SR strategy for self-sufficient resolution enhancement on ToF camera's output images

- amplitude image and
- depth map using data accquired with only a single 3D PMD sensor.

2. State-of-the-art: dependence on additional HR intensity data

Single image super-resolution (SISR)

- currently, artificial intelligence learning-based algorithms reach the highest image quality in SISR results
- Convolutional Neural Networks (CNNs) learn either a per-pixel loss or a perceptual loss between its output and a ground truth image

Shallow → Deep

2014, SRCNN [1] [30.49 dB] → 2017, DRRN [2] [31.68 dB]

2014, SRCNN [1] [30.49 dB] → 2016, Perc. Loss [3] [27.09 dB]

2016, Perc. Loss [3] [27.09 dB] → 2017, **SRGAN**/SRResNet [4] [29.4 dB/ 32.05 dB] → 2017, **MDSR**/EDSR [5] [32.6 dB/ 32.62 dB]

2016, Perc. Loss [3] [27.09 dB] → 2017, **ENet** [6] [31.74 dB]

[dB]: ⌀ PSNR on *Set5* [7] @ 4x-scaling

Map on CNN based SISR methods (excerpt).

Guided depth map super-resolution (GDMSR)

- state-of-the-art methods for GDMSR are mostly optimization-based or learning-based algorithms
- GDMSR requires an additional high-resolution intensity image for guidance

Optimization-based: 2005, MRF [8] [RMSE = 2.24] → 2011, Park et al. [9] [RMSE = 1.82] → 2013, Ferstl et al. [10] [RMSE = 1.29] → 2017, Jung et al. [11] [RMSE = 1.26]

RMSE on Middlebury 2005 *Art disparity map* [12, 13] @ 4x-scaling

Learning-based: 2016, FCN-PDN [14] Deep CNN + variational optimization → 2016, Song et al. [15] Deep CNN + depth statistics and color-depth correlation → 2016, MSG-Net [16] Multi-Scale Guided CNN

Map on GDMSR methods (excerpt).

3. Contribution: independence from further sensor data

Step 1: Superresolve PMD sensor's low-resolution (LR) intensity image using *ENet-PAT* [6] CNN

Perceptual loss

$$\mathcal{L}_P = \|\Phi(I_{est}) - \Phi(I_{HR})\|_2^2$$

$$\Phi(I) = feature\ map\ of\ image\ I$$

→ Euclidean loss optimization on feature maps

Adversarial training

$$\mathcal{L}_A = -\log\Big(D\big(G(z)\big)\Big)$$

$$D = discriminative\ network$$
$$G(z) = generated\ sample$$

→ discriminative network trains mapping from LR images to HR images

Texture matching loss

$$\mathcal{L}_T = \Big\|G\big(\Phi(I_{est})\big) - G\big(\Phi(I_{HR})\big)\Big\|_2^2$$

$$G\big(\Phi(I)\big) = gram\ matrix\ of\ feature\ map\ \Phi(I)$$

→ enforces locally similar textures between SR result and HR ground truth

Step 2: Superresolve LR depth map using an *intensity guided SR* algorithm [11] with the SISR results from step 1

Edge-aware weight

$$W_{ID,p} = \begin{cases} f(W_I)\cdot\big(W_{D,p}+\varepsilon\big)^{\beta}, & \big(W_{I,p}-T_I\big)\cdot\big(W_{D,p}-T_D\big) < 0 \\ \big(W_{I,p}+\varepsilon\big)^{\alpha}\cdot\big(W_{D,p}+\varepsilon\big)^{\beta}, & otherwise \end{cases}$$

$W_{ID,p} = magnitude\ function$ $\quad W_D, W_I = magnitude\ of\ depth/intensity\ gradient$

$f(W_I) = constant\ for\ magnitude\ image\ W_I, balances\ W_{ID}\ for\ different\ cases\ of\ W_I$

$p = for\ a\ pixel$ $\quad \alpha,\beta,\varepsilon = positive\ constants$ $\quad T_I, T_D = pre-defined\ thresholds$

→ controls L_0 gradient regularization term to preserve edges and remove edge blurring and texture copying artifacts

Weighted L_0 gradient minimization

$$\min_{D_H}\left\{\sum_p \big(D_{H,p}-D_p\big)^2 + \frac{\lambda}{W_{ID,p}}\cdot H(\nabla D_{H,p})\right\}$$

$D_H = HR\ depth\ reconstruction$ $\quad D = HR\ depth\ estimation$

$\nabla D_{H,p} = gradient\ of\ D_H\ for\ a\ pixel\ p$ $\quad p = for\ a\ pixel$

$\lambda = positiv\ constant$ $\quad H\big(\nabla D_{H,p}\big) = binary\ function$

→ combines the original L_0 gradient minimization and the magnitude function $W_{ID,p}$

4. Results

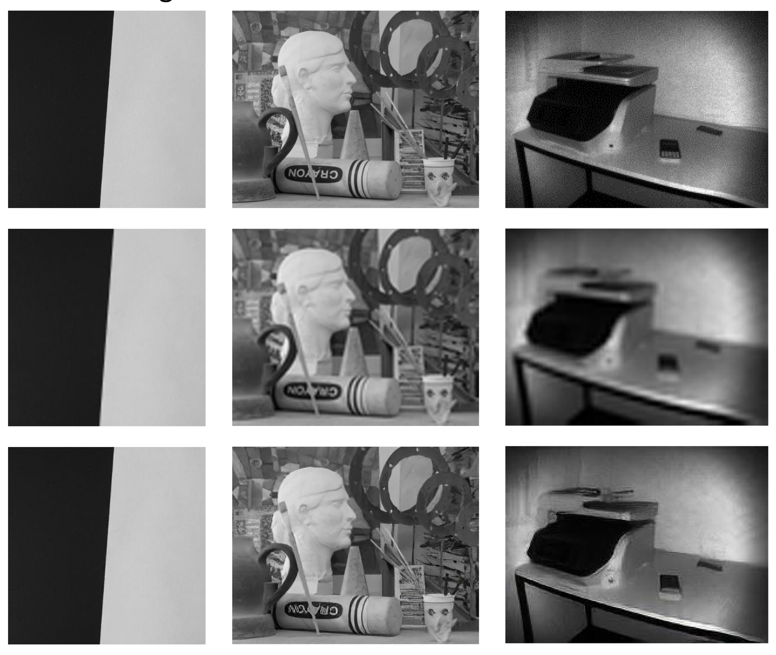
4.1 SISR results on intensity images

Slanted edge | Art | PMDtec PicoFlexx

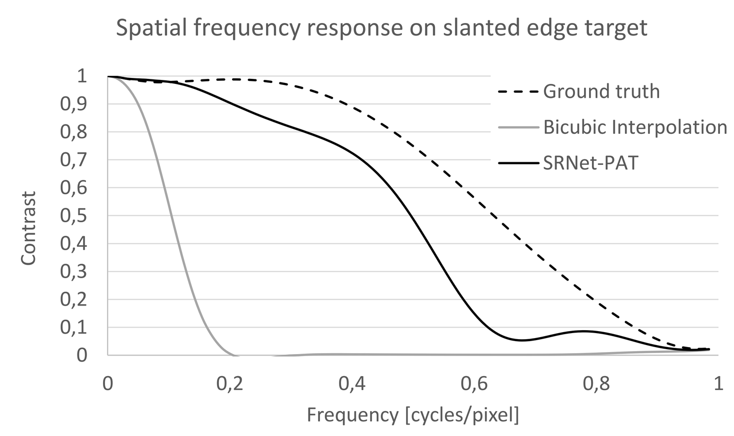
Ground truth

Bicubic Interpolation

ENet-PAT



SISR results on *Slanted edge* target	MTF50 [c/p]	PSNR [dB]	RMSE [a.u.]
Ground truth	0.632	Infinite	0
Bicubic interpolation	0.105	34.37528	4.87277
ENet-PAT SR result[1]	0.497	41.46295	2.15473



Spatial frequency response is measured with MTFMapper [17].

Intensity image [PSNR in dB / RMSE a.u.]	Synthetic: Middlebury 2005 dataset [12, 13]			Real data:
	Art	Books	Moebius	*PMDtec PicoFlexx*
Nearest neighbor interp.	23.78783 / 16.48726	24.05217 / 15.99306	26.39474 / 12.21247	25.50144 / 13.53532
Bicubic interpolation	25.32300 / 13.81626	25.48541 / 13.56031	27.82000 / 10.36430	26.90039 / 11.52181
ENet-PAT SR result[1]	26.63320 / 11.88174	26.57402 / 11.96297	28.10751 / 10.02685	29.52067 / 8.52131

[1]Pre-trained reference implementation of ENet-PAT [6] for magnification ratio of 4

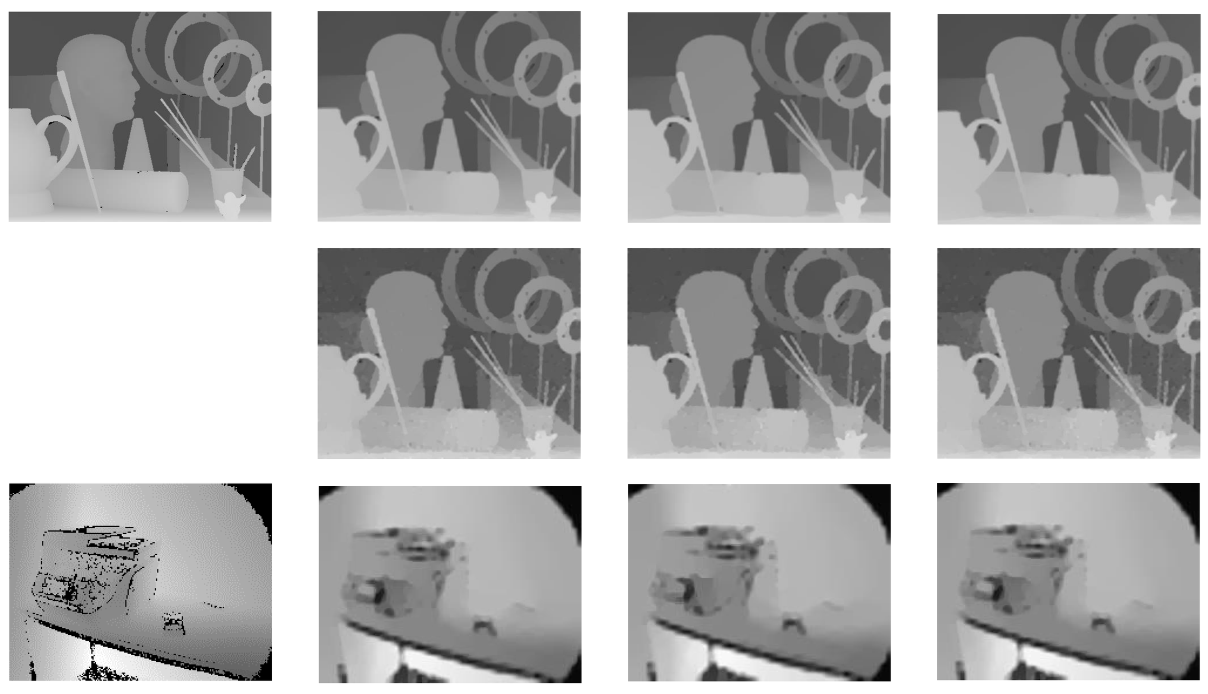
4.2 GDMSR results on depth maps

Ground truth depth | Ground truth guided | Bicubic guided | ENet-PAT guided

Art (noise-free)

Art (noisy[2])

Real data: PMDtec PicoFlexx



Depth image [PSNR in dB / RMSE a.u.]		Synthetic: Middlebury 2005 dataset [12, 13]			Real data:
		Art	Books	Moebius	*PMDtec PicoFlexx*
Noise-free LR depth maps	Ground truth guided	29.12408 / 8.91941	28.93634 / 9.11429	30.11049 / 7.96188	
	Nearest neighbor guided	27.85658 / 10.32075	28.49669 / 9.58751	29.37705 / 8.66339	
	Bicubic guided	28.99904 / 9.04874	28.75584 / 9.30568	29.76623 / 8.28378	
	ENet-PAT guided	**29.11472 / 8.92903**	**28.97976 / 9.06884**	**30.04139 / 8.02547**	
Noisy[2] LR depth maps	Ground truth guided	28.92745 / 9.12363	28.90592 / 9.14627	30.00915 / 8.05532	18.17230 / 31.47205
	Nearest neighbor guided	27.88768 / 10.28387	28.56610 / 9.51120	29.39432 / 8.64617	17.85728 / 32.63442
	Bicubic guided	28.78891 / 9.27032	28.74017 / 9.32249	29.82468 / 8.22822	18.10915 / 31.70173
	ENet-PAT guided	**28.78996 / 9.26920**	**28.95732 / 9.09230**	**29.91369 / 8.14434**	**18.16454 / 31.50019**

[2]Synthetic images are imposed by additive white Gaussian noise with variance $\sigma^2 = 0.001$

5. Conclusions & Outlook

SISR

- good performance on simple *slanted edge* target
  - → SR result reaches nearly 78 % of ground truth's MTF50-value
  - → ENet-PAT's PSNR value is around 1.2 times higher than bicubic interpolation ones
- less image quality on *Art* and *PMDtec PicoFlexx* images
  - → fine details are missing
  - → ENet-PAT's PSNR is only 1.05 times and 1.1 times higher than bicubic interpolation ones for *Art* and *PMDtec PicoFlexx*, respectivley

GDMSR

- image quality is nearly the same for ground truth guided and SISR guided SR depth map results
- moderate overall performance
  - → even the ground truth guided SR results on noise-free LR inputs look blurry
  - → image quality is worse for real data and noisy synthetic images

Outlook

- increase image quality of SISR results by using own training data
- enhance real depth map's image quality by inpainting invalid pixel regions before applying the SR method
- investigate further (learning-based) GDMSR algorithms

References

[1] C. Dong, C. C. Loy, K. He, and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks", in IEEE Transactions on Pattern Analysis and Machine Intelligence 38 (2), (2014)
[2] Y. Tai, J. Yang, and X. Liu, "Image Super-Resolution via Deep Recursive Residual Network", in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (2017)
[3] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual Losses for Real-Time Style Transfer and Super-Resolution", in 14th European Conference on Computer Vision (ECCV), (2016)
[4] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunnignham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network", in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (2017)
[5] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced Deep Residual Networks for Single Image Super-Resolution", in IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), (2017)
[6] M.S.M. Sajjadi, B. Schölkopf, and M. Hirsch, „EnhanceNet: Single Image Super-Resolution through Automated Texture Synthesis" in International Conference on Computer Vision, (2017)
[7] M. Bevilacqua, A. Roumy, C. Guillemot, and M.L. Alberi-Morel, "Low-Complexity Single-Image Super-Resolution based on Nonnegative Neighbor Embedding", in British Machine Vision Conference (BMVC), (2012)
[8] J. Diebel, and S. Thrun, "An Application of Markov Random Fields to Range Sensing", in Proceedings of Conference on Neural Information Processing Systems (NIPS), (2005)
[9] J. Park, H. Kim, Y.W. Tai, M.S. Brown, and I. Kweon, "High Quality Depth Map Upsampling for 3D-TOF Cameras", in International Conference on Computer Vision (ICCV), (2011)
[10] D. Ferstl, C. Reinbacher, R. Ranftl, M. Rüther, and H. Bischof, "Image Guided Depth Upsampling using Anisotropic Total Generalized Variation", in IEEE International Conference on Computer Vision (ICCV), (2013)
[11] C. Jung, S. Yu, and J. Kim, "Intensity-guided edge-preserving depth upsampling through weighted L_0 gradient minimization", in J. Vis. Commun. Image R. 42, (2017)
[12] H. Hirschmüller, and D. Scharstein, "Evaluation of Cost Functions for Stereo Matching", in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (2007)
[13] D. Scharstein, and C. Pal, "Learning Conditional Random Fields for Stereo", in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (2007)
[14] G. Riegler, D. Ferstl, M. Rüther, and H. Bischof, "A Deep Primal-Dual Network for Guided Depth Super-Resolution", in Proceedings of the British Machine Vision Conference (BMVC), (2016)
[15] X. Song, Y. Dai, and X. Qin, "Deep Depth Super-Resolution: Learning Depth Super-Resolution Using Deep Convolutional Neural Network", in 13th Asian Conference on Computer Vision (ACCV), (2016)
[16] T.W. Hui, C.C. Loy, and X. Tang, "Depth Map Super-Resolution by Deep Multi-Scale Guidance", in 14th European Conference on Computer Vision (ECCV), (2016)
[17] F. van den Bergh, "Deferred slanted-edge analysis: a unified approach to spatial frequency response measurement on distorted images and color filter array subsets", in Journal of the Optical Society of America A 35 (3), (2018)

University of Applied Sciences Ravensburg-Weingarten
Faculty T – Technology and Management
Doggenriedstraße
88250 Weingarten, Germany

Laboratory for 3D Camera Technology and Technical Optics
Henrik Lietz, +49 (0) 751/501-9283
henrik.lietz@hs-weingarten.de

DGaO-Proceedings 2018 - http://www.dgao-proceedings.de - ISSN: 1614-8436 - urn:nbn:de:0287-2018-P028-5
eingegangen: 22.05.2018 veröffentlicht: 25.06.2018