Deep Learning Based Cross-Spectral Disparity Estimation for Stereo Imaging (ELI-01 -- Electronic Imaging)

Nils Genser, Andreas Spruck, Jürgen Seiler, and André Kaup

nils.genser@fau.de

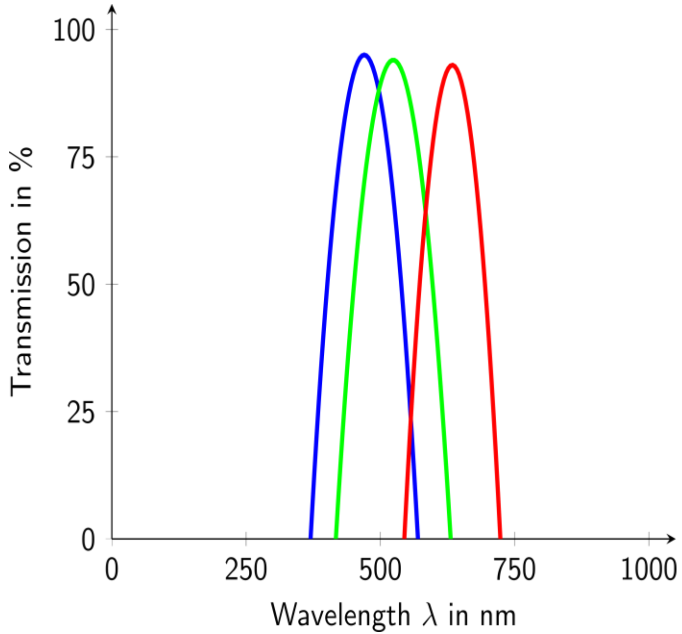
FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

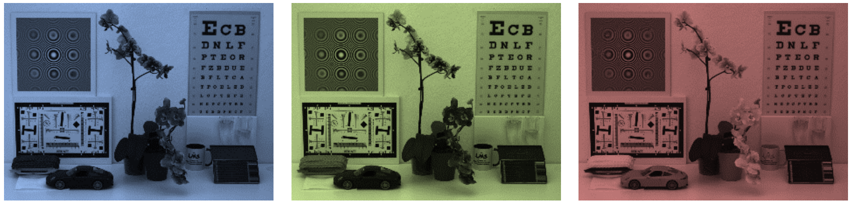
Chair of Multimedia Communications and Signal Processing

LMS

Spectral Imaging

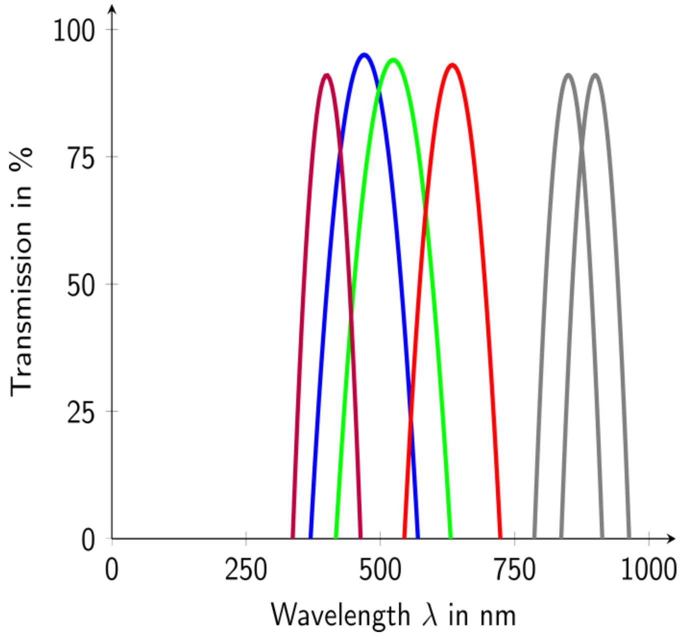
Capture specific light spectra, e.g., blue, green, red

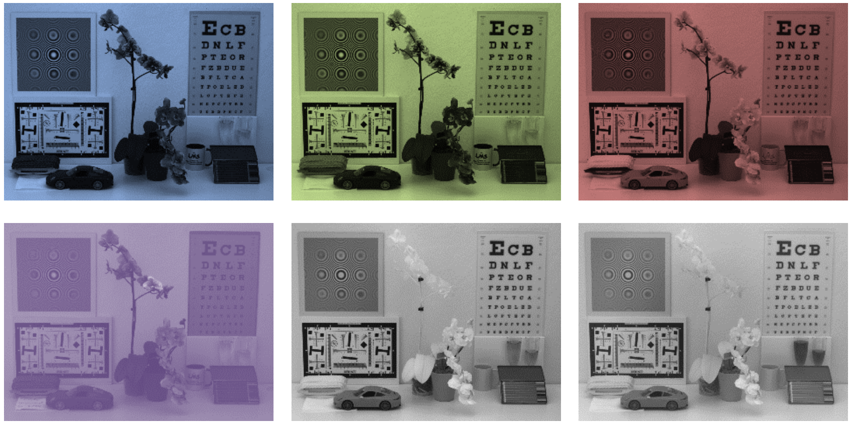




Spectral Imaging

Capture specific light spectra, e.g., blue, green, red, UV, NIR

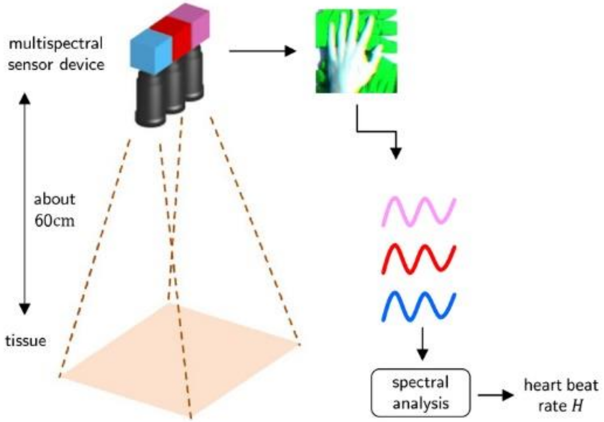




Applications

Spectral imaging:

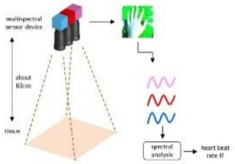
- Healthcare

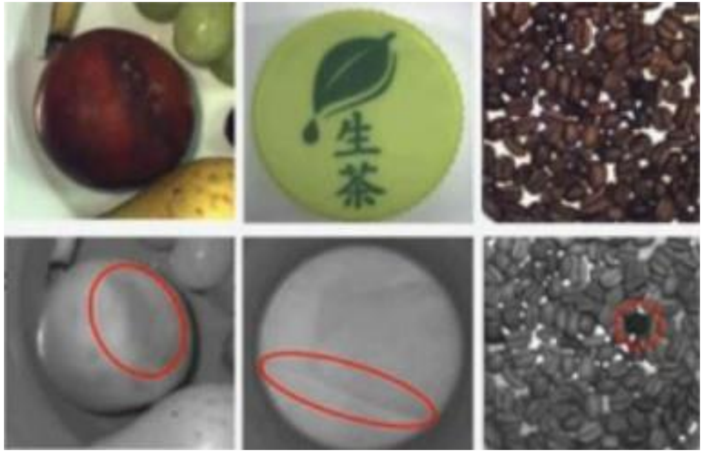


Applications

Spectral imaging:

- Healthcare
- Food quality assessment

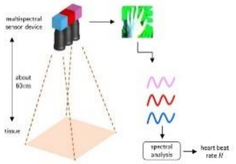


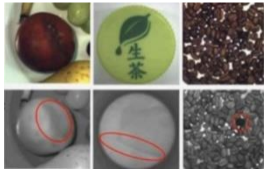


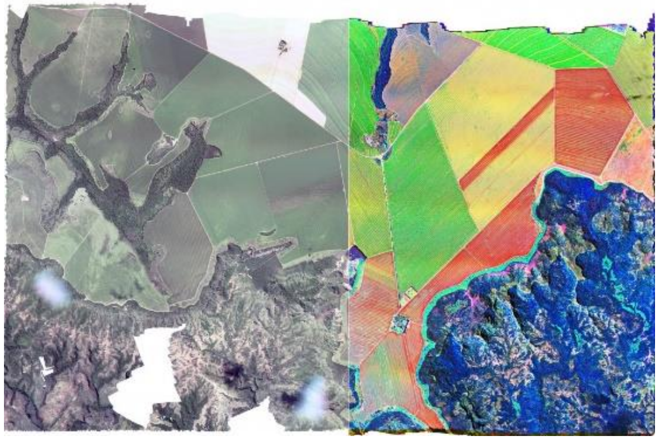
Applications

Spectral imaging:

- Healthcare
- Food quality assessment
- Agriculture



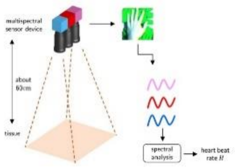


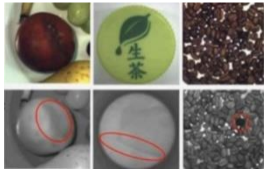


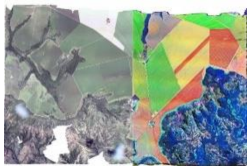
Applications

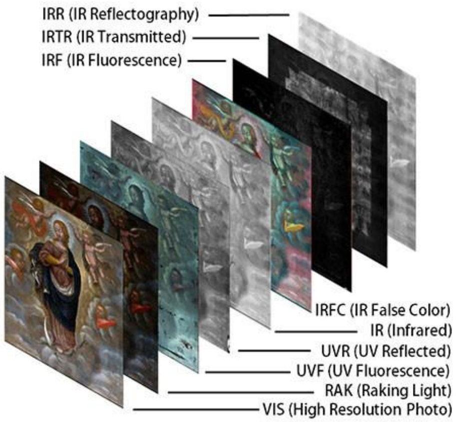
Spectral imaging:

- Healthcare
- Food quality assessment
- Agriculture
- Restoration





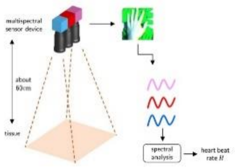


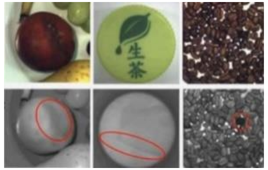


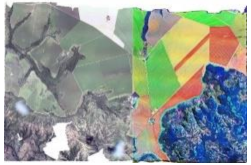
Applications

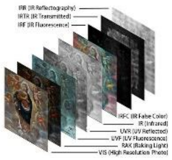
Spectral imaging:

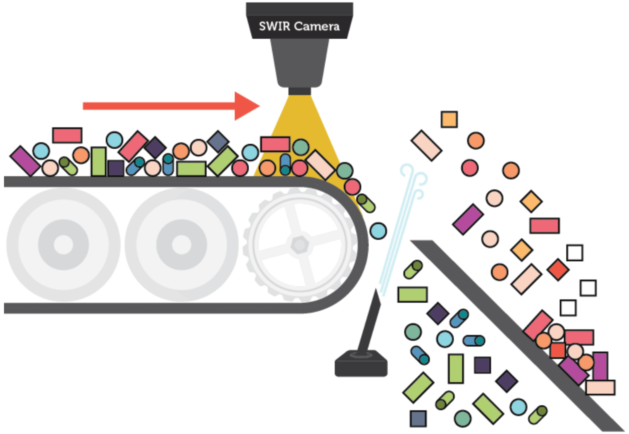
- Healthcare
- Food quality assessment
- Agriculture
- Restoration
- Recycling











Outline

- Motivation
- State-of-the-Art Techniques
- Novel Cross-Spectral Training for Deep Learning
- Evaluation
- Conclusion and Reference Implementation

Stereo Cameras with Different Spectral Filters

Self-manufactured prototype

Stereo Cameras with Different Spectral Filters

Self-manufactured prototype

→

Cheap spectral imager:
- Multi-spectral data
- Depth information

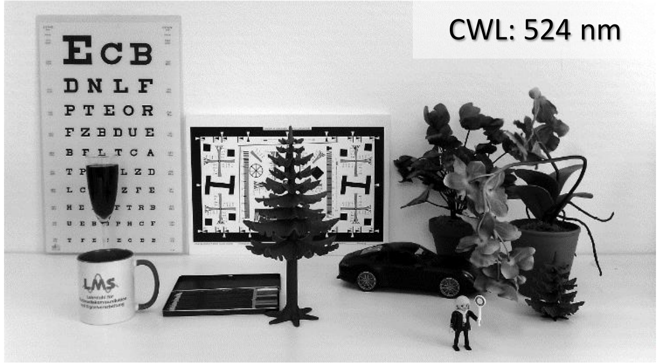
Stereo Cameras with Different Spectral Filters

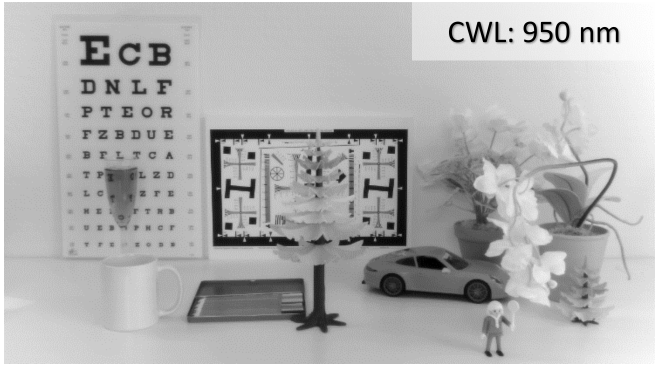
Self-manufactured prototype

→

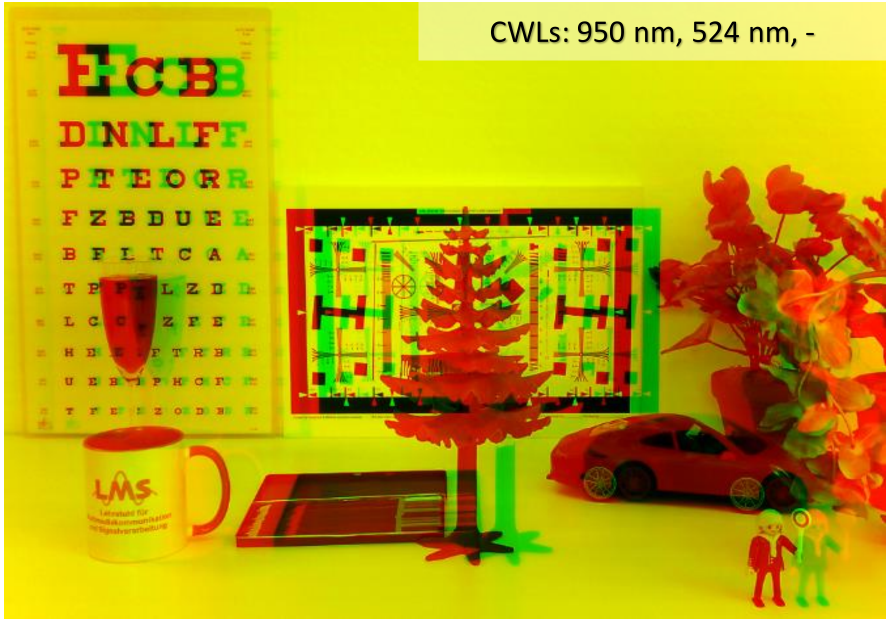
Cheap spectral imager:

- Multi-spectral data
- Depth information





Problem Statement and Challenges



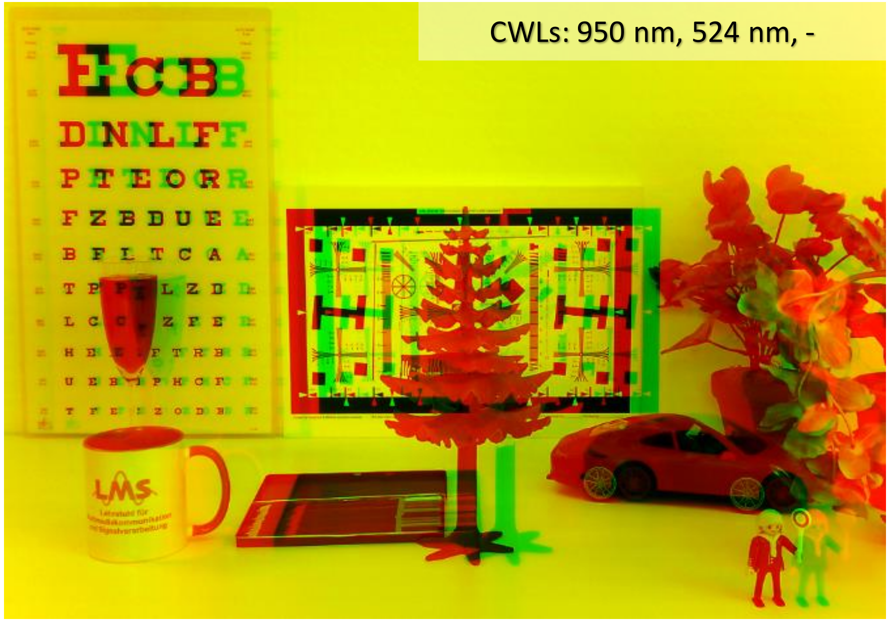
Composed false color image

Problem Statement:

- Recordings at different positions
- Different content (due to spectral filters) at different positions

➢ Pixel-wise compensation (disparity)

Problem Statement and Challenges



Composed false color image

Problem Statement:

- Recordings at different positions
- Different content (due to spectral filters) at different positions
- ➢ Pixel-wise compensation (disparity)

Conventional stereo imaging

- Recordings at different positions
- Same content

State-of-the-Art Cross-Spectral Disparity Estimation

Cross-Spectral Disparity Estimation:

- Design of structural similarity cost functions, e.g., **[1], [2], [3]**
- Cost aggregation, e.g., semi-global matching **[4]**

[1] T. Rukkanchanunt, T. Shibata, M. Tanaka, and M. Okutomi, “Disparity map estimation from cross-modal stereo,” in Proc. IEEE Global Conference on Signal and Information Processing (GlobalSIP), Nov. 2018, pp. 988–992.
[2] O. Zeglazi, M. Rziza, A. Amine, and C. Demonceaux, “Accurate dense stereo matching for road scenes,” in Proc. IEEE International Conference on Image Processing (ICIP), Sept. 2017, pp. 720–724.
[3] S. Mattoccia, F. Tombari, and L. Di Stefano, “Reliable rejection of mismatching candidates for efficient zncc template matching,” in Proc. IEEE International Conference on Image Processing (ICIP), Oct. 2008, pp. 849–852.
[4] H. Hirschmüller, “Accurate and efficient stereo processing by semi-global matching and mutual information,” in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), June 2005, vol. 2, pp. 807–814.
[5] J. Chang and Y. Chen, “Pyramid stereo matching network,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018, pp. 5410–5418.
[6] Z. Yin, T. Darrell, and F. Yu, “Hierarchical discrete distribution decomposition for match density estimation,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2019, pp. 6037–6046.

State-of-the-Art Cross-Spectral Disparity Estimation

Cross-Spectral Disparity Estimation:

- Design of structural similarity cost functions, e.g., **[1], [2], [3]**
- Cost aggregation, e.g., semi-global matching **[4]**

Conventional Deep Learning-Based Disparity Estimation

- Mono-modal registration (same spectral filters, typically RGB) **[5], [6]**
- Fast as well as high-quality disparity estimation
- Outperforms classical template matching approaches

[1] T. Rukkanchanunt, T. Shibata, M. Tanaka, and M. Okutomi, "Disparity map estimation from cross-modal stereo," in Proc. IEEE Global Conference on Signal and Information Processing (GlobalSIP), Nov. 2018, pp. 988–992.
[2] O. Zeglazi, M. Rziza, A. Amine, and C. Demonceaux, "Accurate dense stereo matching for road scenes," in Proc. IEEE International Conference on Image Processing (ICIP), Sept. 2017, pp. 720–724.
[3] S. Mattoccia, F. Tombari, and L. Di Stefano, "Reliable rejection of mismatching candidates for efficient zncc template matching," in Proc. IEEE International Conference on Image Processing (ICIP), Oct. 2008, pp. 849–852.
[4] H. Hirschmüller, "Accurate and efficient stereo processing by semi-global matching and mutual information," in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), June 2005, vol. 2, pp. 807–814.
[5] J. Chang and Y. Chen, "Pyramid stereo matching network," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018, pp. 5410–5418.
[6] Z. Yin, T. Darrell, and F. Yu, "Hierarchical discrete distribution decomposition for match density estimation," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2019, pp. 6037–6046.

Cross-Spectral training for Deep Learning (CSDL)

THE NOVEL SOLUTION APPROACH

Motivation

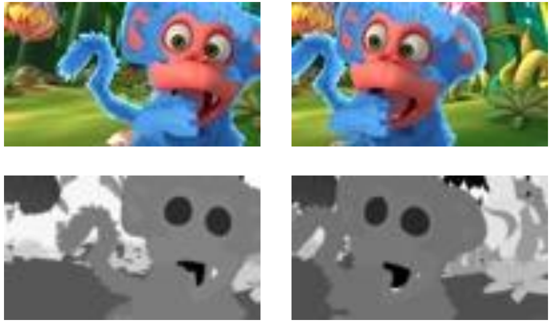
Color stereo training sets:

- Publicly available
- E.g., Flying Things 3D, Driving, Monka **[7]**

➢ RGB stereo matching algorithms trainable

Color images

Disparity maps



[7] N. Mayer, E. Ilg, P. Häusser, P. Fischer, D. Cremers, A. Dosovitskiy, and T. Brox, “A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation,” in Proc. IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), June 2016, pp. 4040–4048.

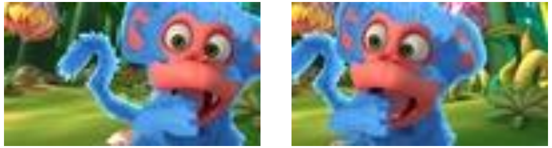
Motivation

Color stereo training sets:

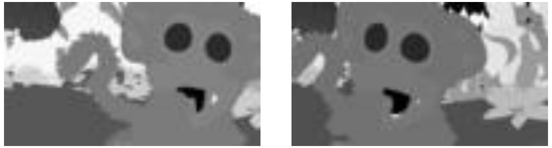
- Publicly available
- E.g., Flying Things 3D, Driving, Monka **[7]**

➢ RGB stereo matching algorithms trainable

Color images



Disparity maps



Cross-spectral stereo training sets:

No data available

[7] N. Mayer, E. Ilg, P. Häusser, P. Fischer, D. Cremers, A. Dosovitskiy, and T. Brox, "A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation," in Proc. IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), June 2016, pp. 4040–4048.

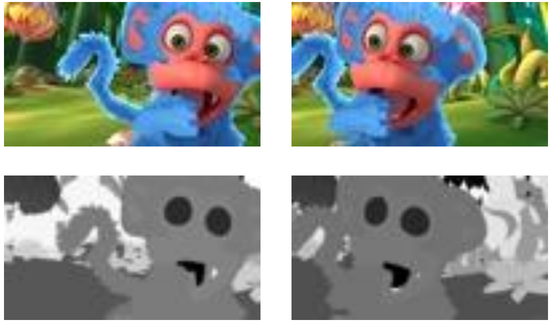
Motivation

Color stereo training sets:

- Publicly available
- E.g., Flying Things 3D, Driving, Monka **[7]**
- ➢ RGB stereo matching algorithms trainable

Color images

Disparity maps



Cross-spectral stereo training sets:

No data available

How to train cross-spectral stereo matching?

➢ **Lack of deep learning-based cross-spectral disparity estimation algorithms!**

[7] N. Mayer, E. Ilg, P. Häusser, P. Fischer, D. Cremers, A. Dosovitskiy, and T. Brox, “A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation,” in Proc. IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), June 2016, pp. 4040–4048.

Fundamentals

Image: $\boldsymbol{I}_c = \int_0^\infty s(\lambda)o(\lambda)l(\lambda)f_c(\lambda)r(\lambda)\mathrm{d}\lambda$

$s(\lambda)$: spectrum light source

$o(\lambda)$: spectral object reflectance

$l(\lambda)$: camera lens transmission

$f_c(\lambda)$: spectral filter transmission

$r(\lambda)$: spectral camera response

Fundamentals

Image: $\boldsymbol{I}_c = \int_0^\infty s(\lambda)o(\lambda)l(\lambda)f_c(\lambda)r(\lambda)\mathrm{d}\lambda$

$s(\lambda)$: spectrum light source

$o(\lambda)$: spectral object reflectance

$l(\lambda)$: camera lens transmission

$f_c(\lambda)$: spectral filter transmission

$r(\lambda)$: spectral camera response

Conclusion:

- For same setup, only spectral filter relevant
- Further filter simulation possible:
  - ➢ Use RGB data sets
  - ➢ Compose known intensity images
  - ➢ Obtain synthesized cross-spectral training set

Fundamentals

Image: $\boldsymbol{I}_c = \int_0^\infty s(\lambda)o(\lambda)l(\lambda)f_c(\lambda)r(\lambda)\mathrm{d}\lambda$

$s(\lambda)$: spectrum light source

$o(\lambda)$: spectral object reflectance

$l(\lambda)$: camera lens transmission

$f_c(\lambda)$: spectral filter transmission

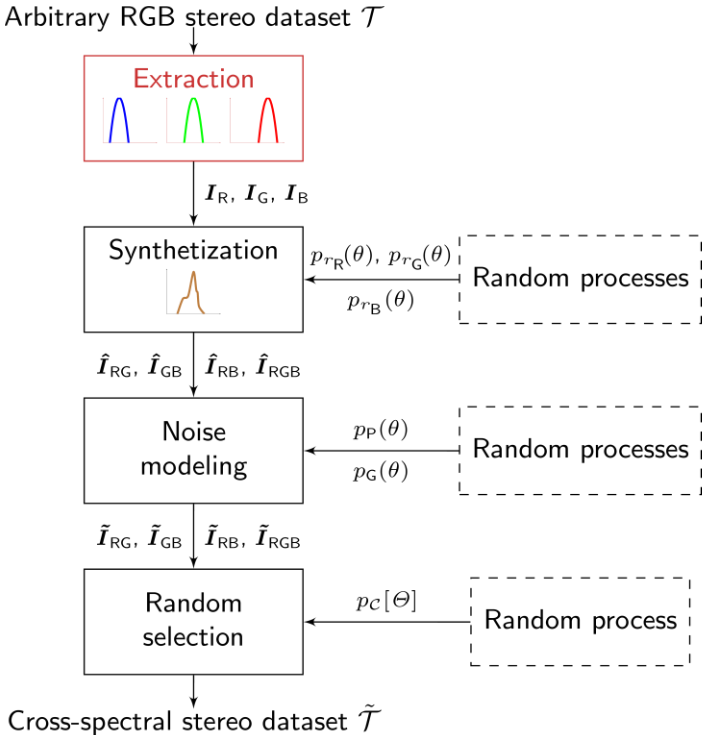
$r(\lambda)$: spectral camera response

Conclusion:

- For same setup, only spectral filter relevant
- Further filter simulation possible:
  - Use RGB data sets
  - Compose known intensity images
  - Obtain synthesized cross-spectral training set

Proposed Cross-Spectral training for Deep Learning (CSDL)

Extraction

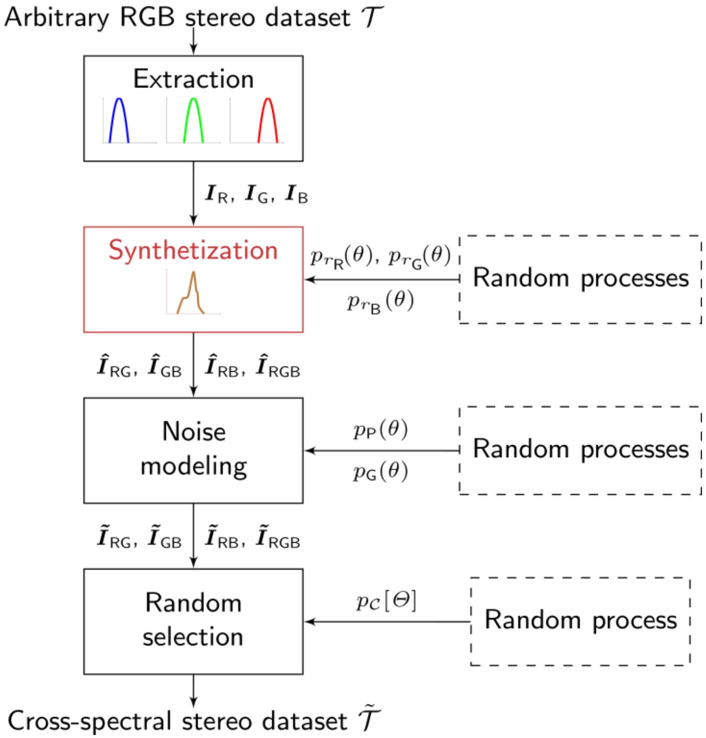


Conventional color stereo training set:

$$\mathcal{T} = \bigcup_{k=0}^{K-1} \bigcup_{c \in \mathcal{C}_{\mathrm{RGB}}} \boldsymbol{I}_{k\,c}, \quad \mathcal{C}_{\mathrm{RGB}} = \{\mathrm{R}, \mathrm{G}, \mathrm{B}\}$$

with set indices k and component indices c

Synthetization



Synthesized channel combinations:

$$\hat{\boldsymbol{I}}_{k\,\mathrm{RG}} = (r_{k\,\mathrm{R}}\boldsymbol{I}_{k\,\mathrm{R}} + r_{k\,\mathrm{G}}\boldsymbol{I}_{k\,\mathrm{G}}) \,/\, (r_{k\,\mathrm{R}} + r_{k\,\mathrm{G}})$$

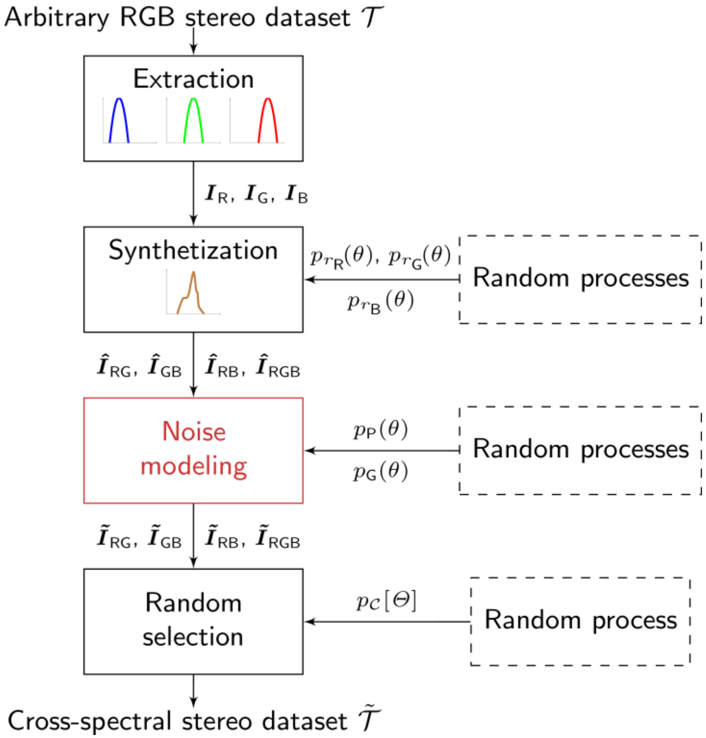
$$\hat{\boldsymbol{I}}_{k\,\mathrm{GB}} = (r_{k\,\mathrm{G}}\boldsymbol{I}_{k\,\mathrm{G}} + r_{k\,\mathrm{B}}\boldsymbol{I}_{k\,\mathrm{B}}) \,/\, (r_{k\,\mathrm{G}} + r_{k\,\mathrm{B}})$$

$$\hat{\boldsymbol{I}}_{k\,\mathrm{RB}} = (r_{k\,\mathrm{R}}\boldsymbol{I}_{k\,\mathrm{R}} + r_{k\,\mathrm{B}}\boldsymbol{I}_{k\,\mathrm{B}}) \,/\, (r_{k\,\mathrm{R}} + r_{k\,\mathrm{B}})$$

$$\hat{\boldsymbol{I}}_{k\,\mathrm{RGB}} = \frac{r_{k\,\mathrm{R}}\boldsymbol{I}_{k\,\mathrm{R}} + r_{k\,\mathrm{G}}\boldsymbol{I}_{k\,\mathrm{G}} + r_{k\,\mathrm{B}}\boldsymbol{I}_{k\,\mathrm{B}}}{r_{k\,\mathrm{R}} + r_{k\,\mathrm{G}} + r_{k\,\mathrm{B}}}$$

with uniformly distributed random variables $r_{k\,\mathrm{R}}, r_{k\,\mathrm{G}}, r_{k\,\mathrm{B}}$ and probability density functions $p_{r_{\mathrm{R}}}(\theta), p_{r_{\mathrm{G}}}(\theta), p_{r_{\mathrm{B}}}(\theta)$

Noise Modeling



Challenges of practical applications: Noise

- Narrowband spectral filters and limited light sources
- ➢ Electronic circuit noise

$$p_{\mathrm{G}}(\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{\theta^2}{2\sigma^2}}, \text{ with } \sigma = 25$$

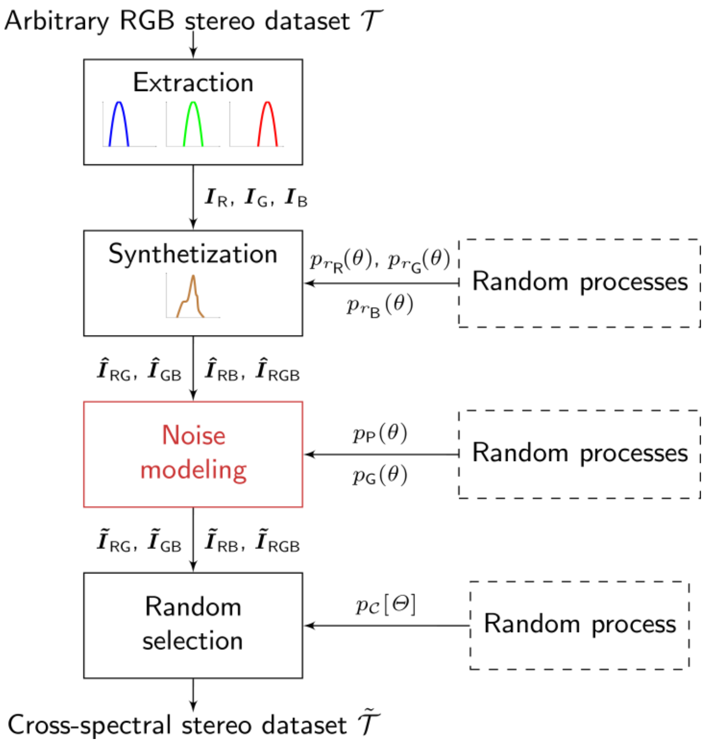
- ➢ Photon shot noise

$$p_{\mathrm{P}}(\theta) = \frac{\Lambda^\theta}{\theta!} \mathrm{e}^{-\Lambda}, \text{ with } \Lambda = 10000$$

- ➢ Amount of photons

$$\gamma = \frac{r_{k\,\mathrm{R}} + r_{k\,\mathrm{G}} + r_{k\,\mathrm{B}}}{3}$$

Noise Modeling



Challenges of practical applications: Noise

- Narrowband spectral filters and limited light sources
- ➢ Electronic circuit noise

$$p_{\mathsf{G}}(\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{\theta^2}{2\sigma^2}}, \text{ with } \sigma = 25$$

- ➢ Photon shot noise

$$p_{\mathsf{P}}(\theta) = \frac{\Lambda^\theta}{\theta!} \mathrm{e}^{-\Lambda}, \text{ with } \Lambda = 10000$$

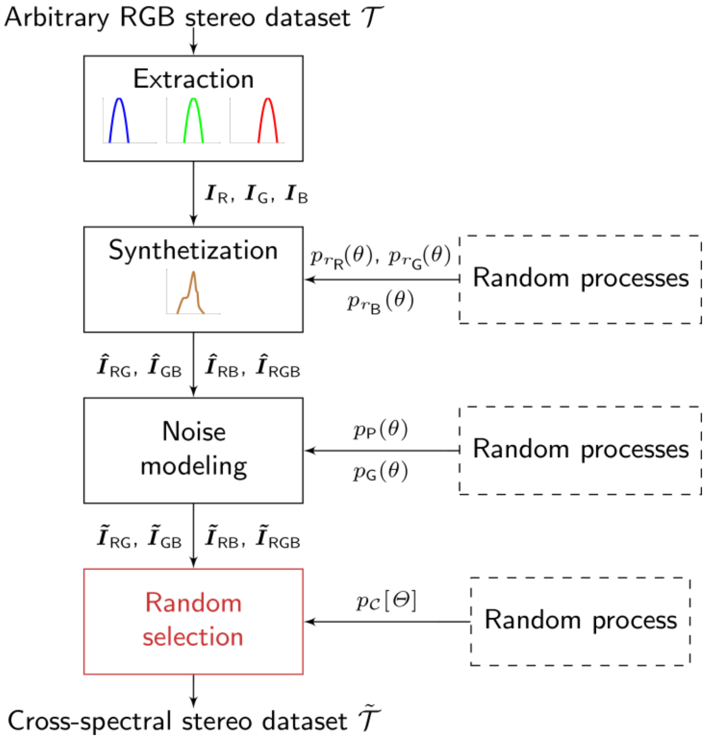
- ➢ Amount of photons

$$\gamma = \frac{r_{k\,\mathsf{R}} + r_{k\,\mathsf{G}} + r_{k\,\mathsf{B}}}{3}$$

Noisy image:

$$\tilde{\boldsymbol{I}}_{k\,c\,x,y} = \max\left(0, \min\left(\frac{p_{\mathsf{P}}(\hat{\boldsymbol{I}}_{k\,c\,x,y}\gamma\Lambda) + p_{\mathsf{G}}(\hat{\boldsymbol{I}}_{k\,c\,x,y})}{\gamma\Lambda}, 1\right)\right)$$

Random Selection



Possible synthesized components:

$$\mathcal{C} = \{\mathrm{R}, \mathrm{G}, \mathrm{B}, \mathrm{RG}, \mathrm{GB}, \mathrm{RB}, \mathrm{RGB}\}$$

Synthesized training set:

$$\tilde{\mathcal{T}} = \bigcup_{k=0}^{K-1} \bigcup_{c \in \hat{\mathcal{C}}} \tilde{I}_{k\ c}$$

with set $\hat{\mathcal{C}}$ holding three randomly selected candidates from $\mathcal{C}$

- RGB and synthesized training set of same size

Evaluation

PERFORMANCE ANALYSIS

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

Objective Evaluation I

Training: Synthetic FlyingThings3D data set **[7]** → RGB and generated CSDL images

Evaluation: Natural Middlebury images **[8]**

Metrics: Bad Matched Pixels (BMP), End-Point Error (EPE)

[7] N. Mayer, E. Ilg, P. Häusser, P. Fischer, D. Cremers, A. Dosovitskiy, and T. Brox, "A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation," in Proc. IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), June 2016, pp. 4040–4048.

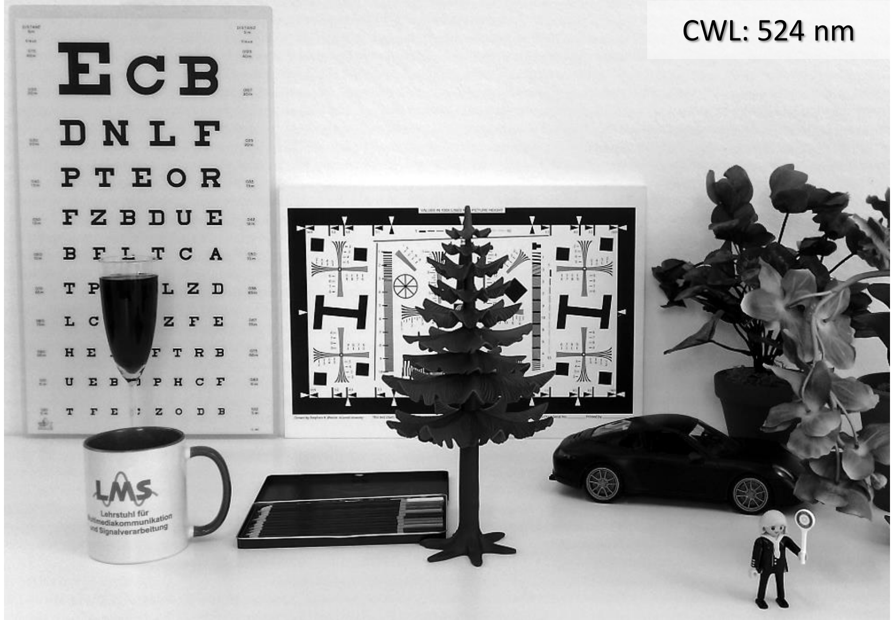
[8] D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Neˇsi´c, X. Wang, and P. Westling, "Highresolution stereo datasets with subpixel-accurate ground truth," in Proc. German Conference on Pattern Recognition (GCPR), Sept. 2014, pp. 31–42.

Objective Evaluation II

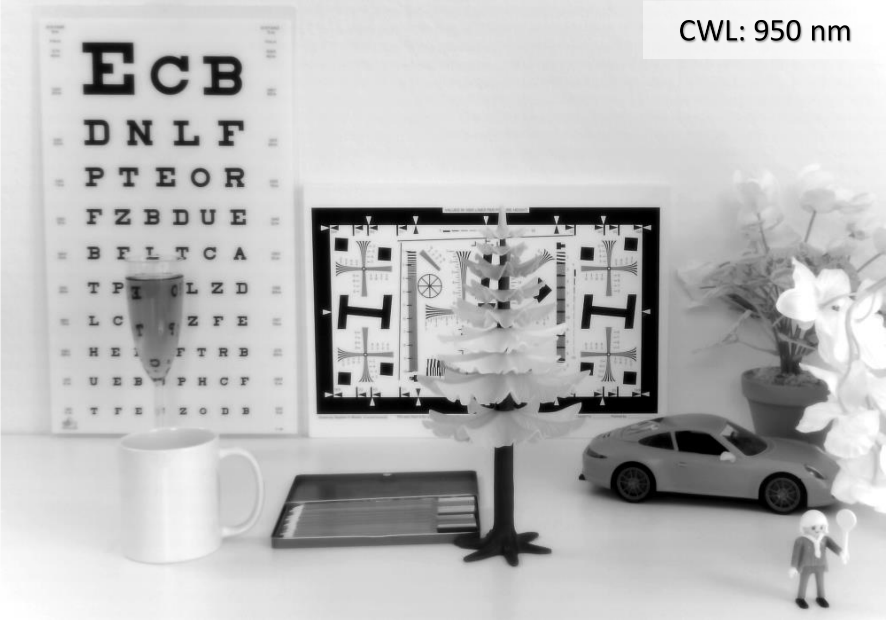
	Registration: red → green			Registration: blue → green			Registration: red → blue		
	BMP-1	BMP-5	EPE	BMP-1	BMP-5	EPE	BMP-1	BMP-5	EPE
Classical: Census [2] + SGM [4]	45.0 %	31.4 %	8.65	45.6 %	35.3 %	8.67	59.0 %	48.2 %	11.48
Classical: ZNCC [3] + SGM [4]	36.9 %	29.7 %	8.21	39.2 %	30.8 %	8.22	50.2 %	39.7 %	10.88
HD^{3}Net: Pretrained RGB [6]	66.2 %	63.6 %	63.91	73.6 %	73.0 %	70.65	79.9 %	78.4 %	79.98
HD^{3}Net: Selftrained RGB	40.8 %	37.6 %	16.64	42.0 %	39.1 %	21.09	54.0 %	48.8 %	23.68
HD^{3}Net: Proposed CSDL	23.4 %	19.7 %	5.73	24.5 %	19.1 %	4.90	29.1 %	24.4 %	6.98
PSMNet: Pretrained RGB [5]	43.2 %	33.5 %	8.83	52.4 %	26.5 %	8.54	58.8 %	41.7 %	12.39
PSMNet: Selftrained RGB	44.6 %	34.7 %	9.05	47.3 %	37.5 %	9.51	58.3 %	50.7 %	14.02
PSMNet: Proposed CSDL	**13.2 %**	**8.7 %**	**1.91**	**14.0 %**	**8.6 %**	**2.29**	**19.8 %**	**10.6 %**	**2.88**

[2] O. Zeglazi, M. Rziza, A. Amine, and C. Demonceaux, "Accurate dense stereo matching for road scenes," in Proc. IEEE International Conference on Image Processing (ICIP), Sept. 2017, pp. 720–724.
[3] S. Mattoccia, F. Tombari, and L. Di Stefano, "Reliable rejection of mismatching candidates for efficient zncc template matching," in Proc. IEEE International Conference on Image Processing (ICIP), Oct. 2008, pp. 849–852.
[4] H. Hirschmüller, "Accurate and efficient stereo processing by semi-global matching and mutual information," in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), June 2005, vol. 2, pp. 807–814.
[5] J. Chang and Y. Chen, "Pyramid stereo matching network," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018, pp. 5410–5418.
[6] Z. Yin, T. Darrell, and F. Yu, "Hierarchical discrete distribution decomposition for match density estimation," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2019, pp. 6037–6046.

Visual Evaluation I

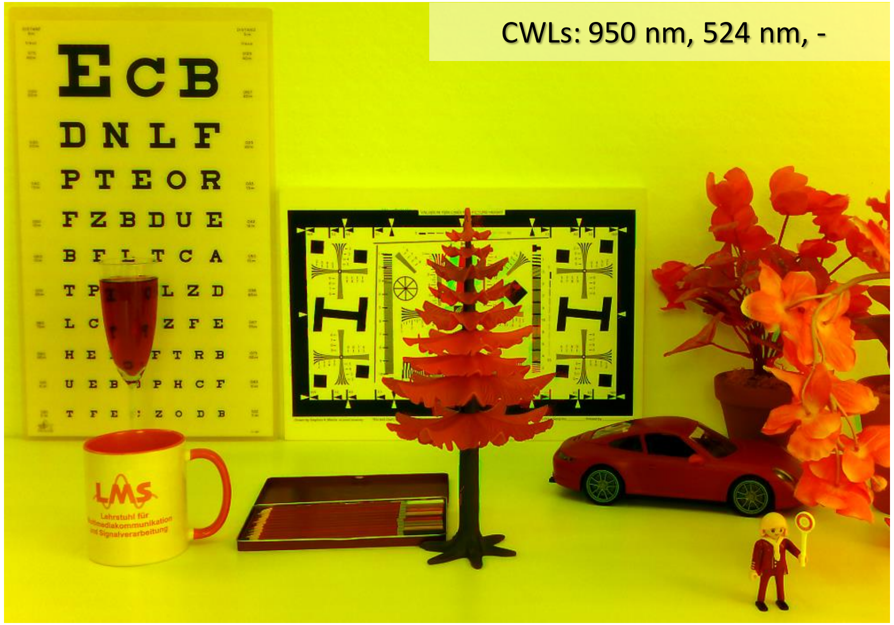


First camera



Second camera

Visual Evaluation II



Registered image (green ← near-infrared, CSDL-trained PSMNet)

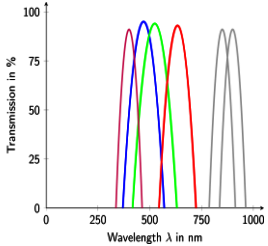
Estimated disparity map

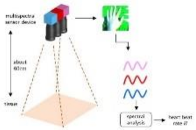
Conclusion

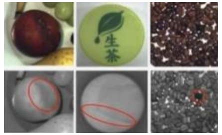
AND REFERENCE IMPLEMENTATION

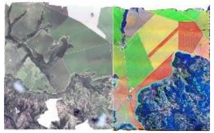
Conclusion

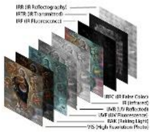
Fundamentals and Applications





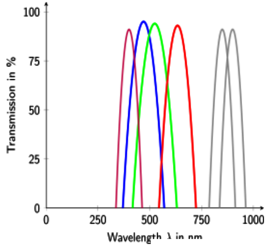


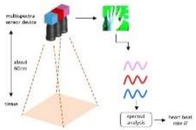


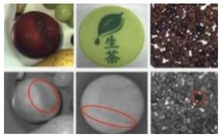


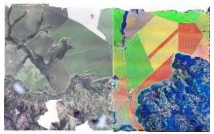
Conclusion

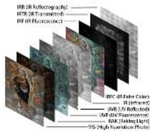
Fundamentals and Applications







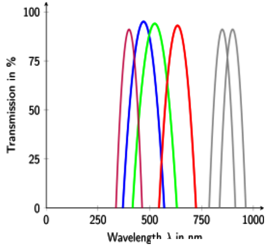


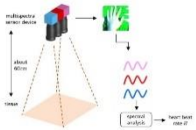


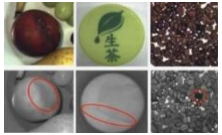
Prototype system

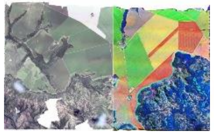
Conclusion

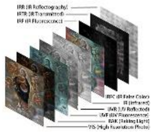
Fundamentals and Applications



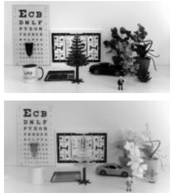




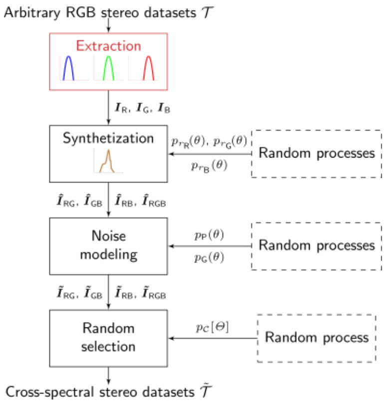




Prototype system

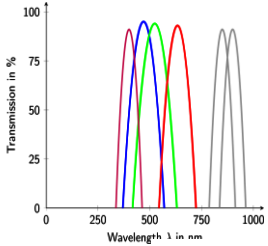


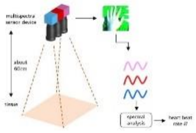
Proposed training algorithm

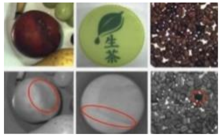


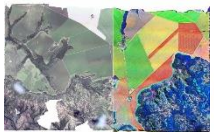
Conclusion

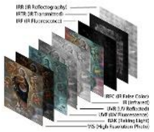
Fundamentals and Applications



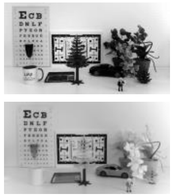




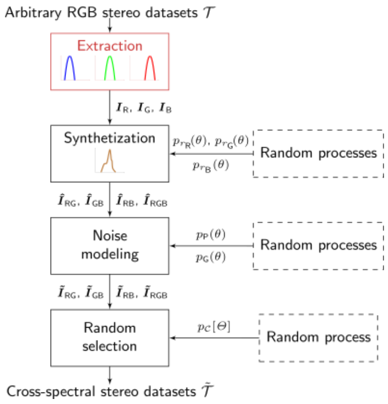




Prototype system



Proposed training algorithm



Objective and visual evaluation

Reference Implementation

CSDL framework :

- Cross-Spectral training for Deep Learning (CSDL)
- *https://gitlab.lms.tf.fau.de/lms/csdl*