Globally Consistent Multi-Label Assignment on the Ray Space of 4D Light Fields

Heidelberg Collaboratory for Image Processing

Sven Wanner, Christoph Straehle and Bastian Goldluecke

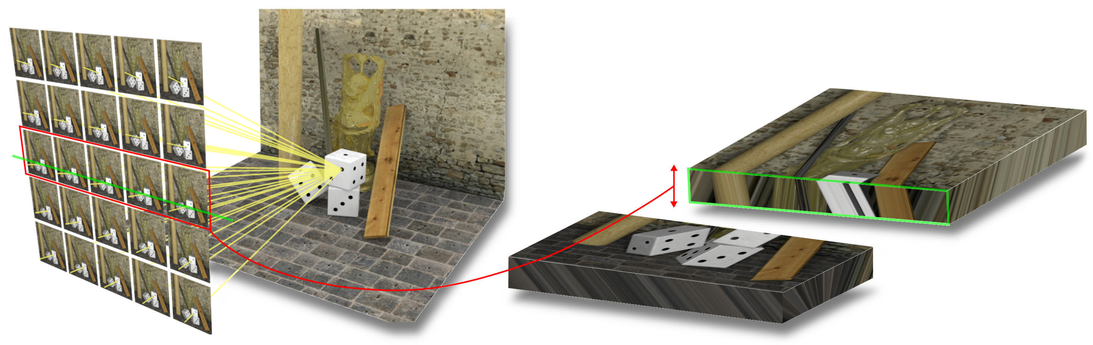
Heidelberg Collaboratory for Image Processing, University of Heidelberg
http://lightfield-analysis.net
e-mail: {sven.wanner,christoph.straehle,bastian.goldluecke}@iwr.uni-heidelberg.de

Contributions

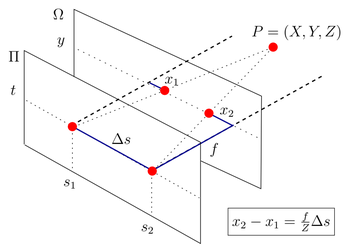
- General segmentation framework using substitutable classifier and optimization components
- Ray space features extend common classifiers to overcome problems of single view classification
- Variational multi-label optimization framework for consistent multi-label assignment on 4D ray space

4D Light Field Parametrization and Epipolar Plane Images (EPIs)

Light field structure



Disparity and epipolar plane images



Light field parametrization

Epipolar plane image (EPI)

Disparity equals local slope

Optimal label assignment on ray space [1]

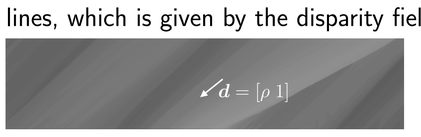
The general inverse problem

Find a vector field $\boldsymbol{U}$ on ray space $\mathcal{R}$ which minimizes

$$\underset{\boldsymbol{U}:\mathcal{R}\to\mathbb{R}^d}{\operatorname{argmin}} \Big\{ \underbrace{J(\boldsymbol{U})}_{\text{convex ray space regularizer}} + \underbrace{F(\boldsymbol{U})}_{\text{convex data term encodes problem}} \Big\}.$$

Regularizer for an epipolar plane image

Regularization needs to be performed in the direction of epipolar lines, which is given by the disparity field ρ:



This can be enforced by using an *anisotropic total variation*

$$J_\rho(\boldsymbol{U}_{y^*,t^*}) := \sum_{i=1}^{n} \int \sqrt{(\nabla U^i_{y^*,t^*})^T D_\rho \nabla U^i_{y^*,t^*}}\ \mathrm{d}(x,s),$$

where the tensor D_ρ encodes the direction information.

Regularizer for ray space

Sum of independent convex regularizers for epipolar plane images in (y,t) and (x,s) coordinates as well as pinhole views in (x,y) coordinates:

$$J_{\lambda\mu}(\boldsymbol{U}) = \mu J_{xs}(\boldsymbol{U}) + \mu J_{yt}(\boldsymbol{U}) + \lambda J_{st}(\boldsymbol{U})$$

$$\text{with } J_{xs}(\boldsymbol{U}) = \int J_\rho(\boldsymbol{U}_{x^*,s^*})\ \mathrm{d}(x^*,s^*)$$

$$J_{yt}(\boldsymbol{U}) = \int J_\rho(\boldsymbol{U}_{y^*,t^*})\ \mathrm{d}(y^*,t^*)$$

$$J_{st}(\boldsymbol{U}) = \int J_V(\boldsymbol{U}_{s^*,t^*})\ \mathrm{d}(s^*,t^*).$$

The variational multi-label problem

Input

- Discrete space of labels Γ
- Arbitrary label assignment cost functions $c_\gamma(R)$

Model

- Set of *binary indicator functions* $u_\gamma : \mathcal{R} \to \{0,1\}$
- Unique assignment: simplex constraint

$$\sum_{\gamma\in\Gamma} u_\gamma = 1.$$

- Encourage consistency with EPI regularizer $J_{\lambda\mu}$.
- Arbitrary spatial regularizer J_V for transition costs.

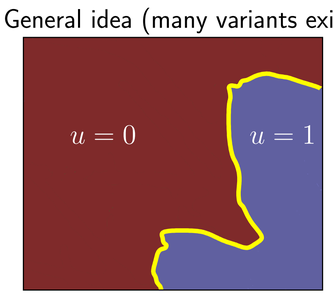
Convex relaxation

$$\underset{\boldsymbol{U}\in\mathcal{C}}{\operatorname{argmin}} \left\{ J_{\lambda\mu}(\boldsymbol{U}) + \sum_{\gamma\in\Gamma} \int_{\mathcal{R}} c_\gamma u_\gamma\ \mathrm{d}(x,y,s,t) \right\},$$

where $\mathcal{C}$ is the convex set of functions $\boldsymbol{U} = (u_\gamma : \mathcal{R} \to [0,1])_{\gamma\in\Gamma}$ which satisfy the simplex constraint.

Spatial regularizer

General idea (many variants exist):



Length of interface equals *total variation*

$$\int_\Omega |Du|\,.$$

Weighted Potts penalizer:

$$J_V(\boldsymbol{U}_{s^*,t^*}) := \frac{1}{2} \sum_{\gamma\in\Gamma} \int_\Omega g\,\left|(Du_\gamma)_{s^*t^*}\right|\ \mathrm{d}(x,y)$$

where g is a spatially varying transition cost.

Optimization

The labeling $\boldsymbol{U}$ is initialized such that the indicator function for the optimal point-wise label is set to one, and zero otherwise. Then we iterate

- data term descent: $U_\lambda \leftarrow U_\lambda - \tau c_\lambda$ for all $\lambda \in \Lambda$,
- EPI regularizer descent ($\boldsymbol{U}_{s^*,t^*}$ constrained to simplex):

$$\boldsymbol{U}_{x^*,s^*} \leftarrow \text{prox}_{\tau\mu J_\rho}(\boldsymbol{U}_{x^*,s^*}) \text{ for all } (x^*,s^*),$$
$$\boldsymbol{U}_{y^*,t^*} \leftarrow \text{prox}_{\tau\mu J_\rho}(\boldsymbol{U}_{y^*,t^*}) \text{ for all } (y^*,t^*),$$

- spatial regularizer descent ($\boldsymbol{U}_{s^*,t^*}$ constrained to simplex):

$$\boldsymbol{U}_{s^*,t^*} \leftarrow \text{prox}_{\tau\lambda J_V}(\boldsymbol{U}_{s^*,t^*}) \text{ for all } (s^*,t^*).$$

The proximation operators prox_J compute subgradient descent steps for the respective 2D regularizer, and enforce the simplex constraint for $\boldsymbol{U}$.

On our system equipped with an nVidia GTX 580 GPU, optimization takes about 1.5 seconds per label in Γ and per million rays in $\mathcal{R}$, i.e. about 5 minutes for our rendered data sets if the result for all views is desired.

IEEE International Conference on Computer Vision and Pattern Recognition 2013, Portland, Oregon, USA.

Local Model

Segmentation model and data term

Features used	Classifier IMG	IMG-D	IMG-GT
RGB value	✓	✓	✓
Intensity standard deviation (in local neighbourhood)	✓	✓	✓
Eigenvalues of Hessian	✓	✓	✓
Laplace operator	✓		
Estimated disparity		✓	
Ground truth disparity			✓

Combination of features used to train a random forest classifier for a local cost function.

Data set	IMG acc	IMG br	IMG-D acc	IMG-D br	IMG-GT acc	IMG-GT br
synthetic data sets						
Buddha	93.5	6.4	96.7	39.6	98.6	43.1
Garden	95.1	54.8	96.7	51.1	96.9	53.3
Papillon	90.8	16.7	96.5	33.1	99.1	73.0
Horses	94.6	15.9	95.3	36.8	98.5	50.9
StillLife	97.8	25.4	98.3	36.1	98.5	39.1
real-world data sets						
UCSD	95.8	8.9	97.0	11.2	-	-
Plenoptic 1	93.7	3.5	94.5	4.4	-	-

Comparison of local labeling accuracy (acc) and boundary recall (br) for all datasets.

Class Probabilities and Regularization Weights

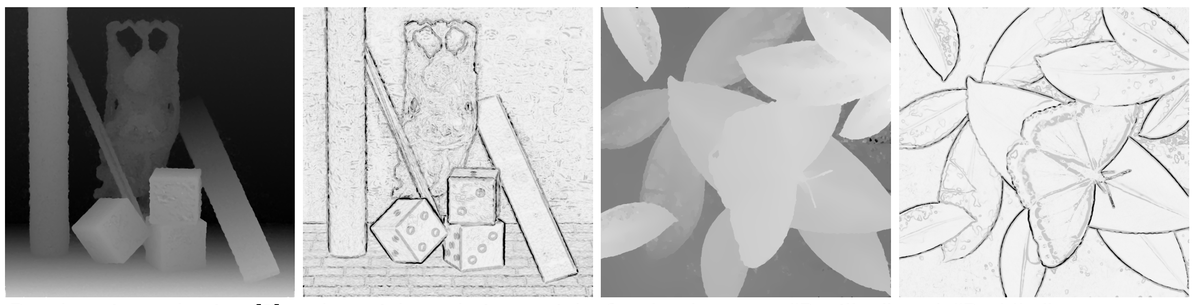
The unary potentials in are initialized with the log-probabilities

$$c_\gamma(R) = -\log p\big(\gamma|\mathbf{v}(R)\big),$$

from the local class probabilities, while the spatial regularization weight g is set to

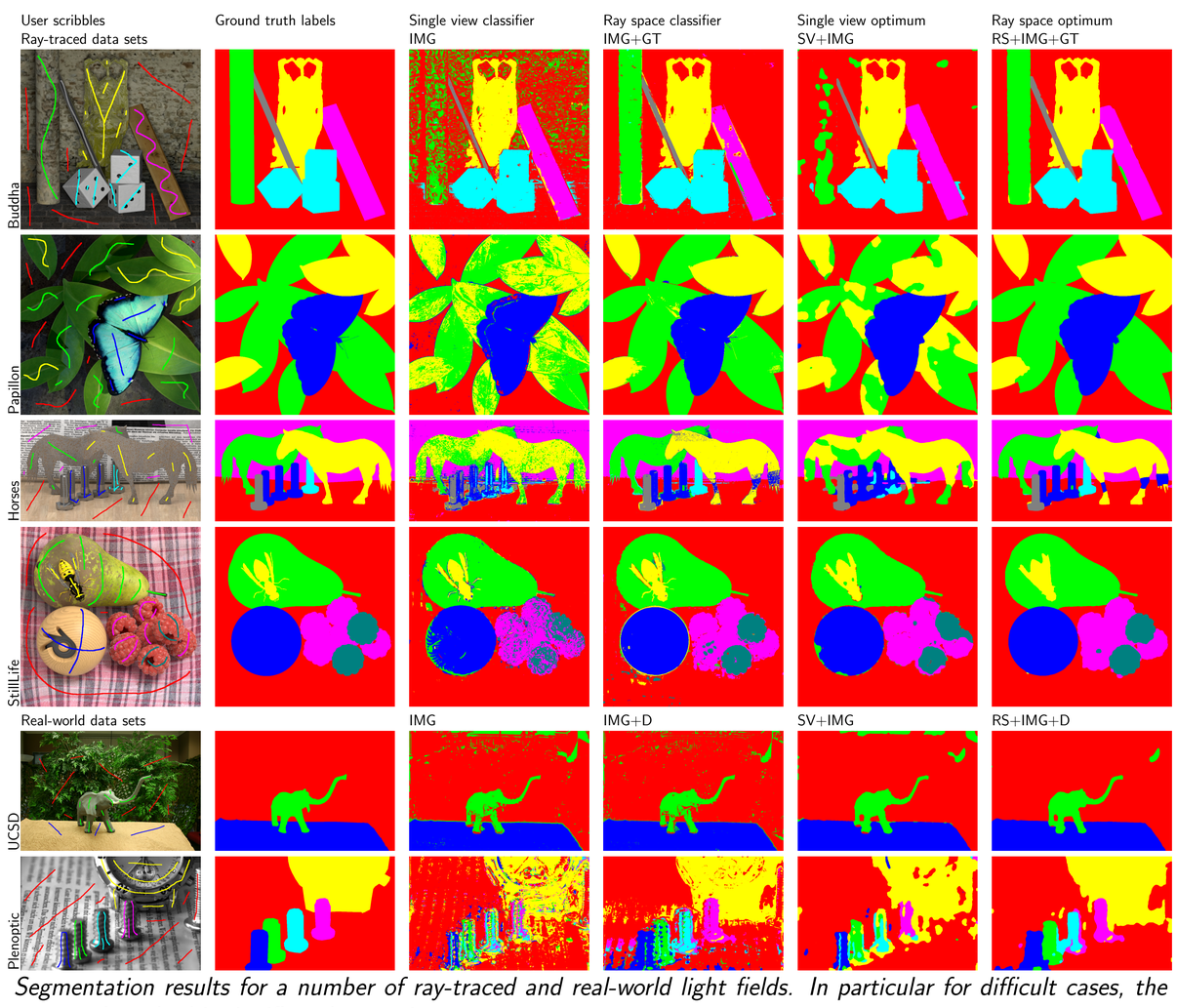
$$g = \max\big\{0, 1 - \big(|\nabla I|_2 - \mathcal{H}(I)\big)\ |\nabla\rho|_2\big\},$$

where I denotes the respective single view image, $\mathcal{H}$ the Harris corner detector, and ρ the disparity field.



Depth estimated using [2] and spatial regularizer weight for the data sets "Papillon" and "Buddha".

Global Optimization (Potts regularizer)



Segmentation results for a number of ray-traced and real-world light fields. In particular for difficult cases, the proposed method is significantly superior.

Optimization	Single view (SV)						Ray space (RS)						Overall improvement	
Classifier	IMG		IMG+D		IMG+GT		IMG		IMG+D		IMG+GT		RS+IMG+D	RS+IMG+GT
	acc	imp	acc	imp	acc	imp	acc	imp	acc	imp	acc	imp	vs. SV+IMG	vs. SV+IMG
Buddha	96.3	43.4	97.5	22.1	99.1	31.2	96.3	43.8	98.8	63.8	99.1	35.5	68.2	76.0
Papillon	92.3	22.4	98.1	46.0	99.3	29.8	93.0	24.6	98.9	68.2	99.5	44.7	84.7	92.8
Horses	96.1	28.4	96.3	21.4	99.0	31.3	96.2	29.5	98.3	64.1	99.1	36.7	56.7	76.2
StillLife	98.8	47.1	98.8	31.0	99.1	41.1	99.0	55.6	98.9	38.5	99.2	45.9	10.1	33.6
UCSD	97.6	44.3	99.1	70	–	–	97.8	48.6	99.3	76.3	–	–	69.9	–
Plenoptic	94.1	34.4	96.1	33.2	–	–	94.5	39.4	96.1	33.9	–	–	34.6	–
Average	96.8	32.2	97.9	36.3	99.2	38.7	96.9	37.1	98.7	55.9	99.4	52.0	41.2	67.1

Relative improvements by global optimization. The quantities in the columns acc *indicate the percentage of correctly labeled pixels, the columns* imp *denote the relative improvement of the optimized compared to the respective raw result from the local classifier. Note that for every single classifier and data set, ray space (RS) optimization achieves a better result than single view (SV) optimization.*

Bibliography

B. Goldluecke and S. Wanner.
The variational structure of disparity and regularization of 4D light fields.
In *Proc. International Conference on Computer Vision and Pattern Recognition*, 2013.

S. Wanner and B. Goldluecke.
Globally consistent depth labeling of 4D light fields.
In *Proc. International Conference on Computer Vision and Pattern Recognition*, pages 41–48, 2012.