Master Thesis Presentation

Comparison of Disparity Algorithms for Stereoscopic Videos

Ben John

May 25, 2016

University of Mannheim, Department of Praktische Informatik IV

Table of contents

Motivation

Applications

- Depth-estimation via camera settings
- Kinect (sunlight)
- 3DTV (remapping)

Foundations

Epipolar geometry

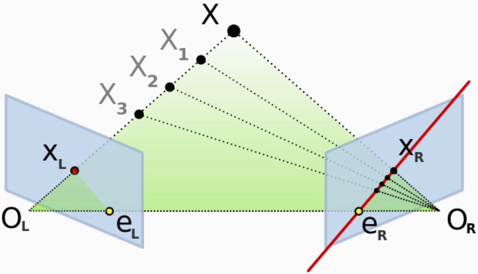


Figure 1: Epipolar geometry[1]

[1] Source (accessed 02/2016): `https://en.wikipedia.org`.

Epipolar geometry

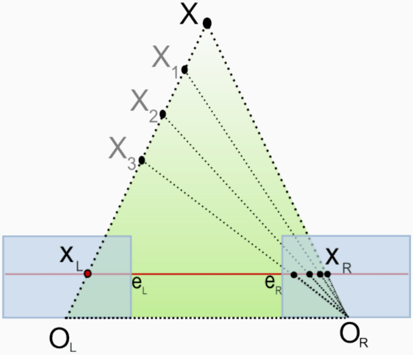


Figure 2: Epipolar geometry after image rectification[2]

[2]Source (accessed 02/2016): `https://en.wikipedia.org`.

Constraints

- Epipolar
- Uniqueness
- Continuity
- Ordering
- Limit
- Lambertian

Classification

- Local methods
  - Area matching
  - Feature matching
- Global methods
  - Dynamic programming
  - Graph cuts
  - Belief propagation

Classification

- Local methods
  - Area matching
  - Feature matching
- Global methods
  - Dynamic programming
  - Graph cuts
  - Belief propagation

Energy function

$$E(d) = E_{data}(d) + \lambda E_{smooth}(d)$$

Taxonomy

1. Computing of matching cost
2. Save values in disparity space image
3. Aggregation of cost values
4. Disparity refinement

Matching cost

- Penalty
- Cost for having dissimilarities
- Optimum $= 0$

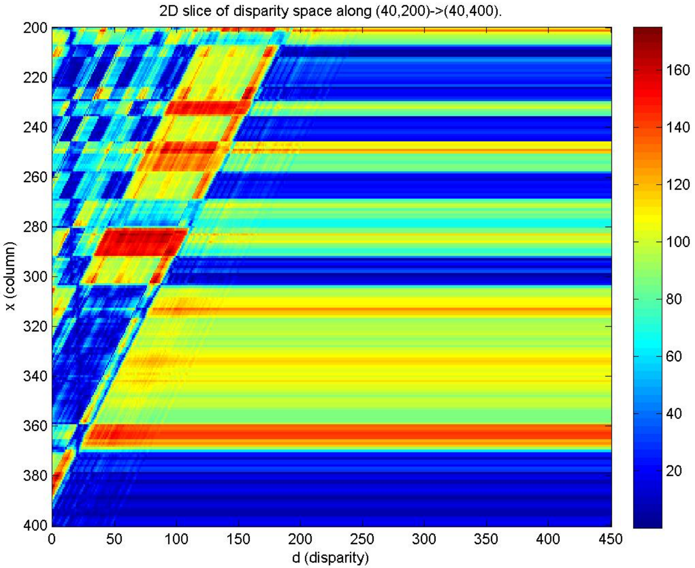
Matching cost

- Penalty
- Cost for having dissimilarities
- Optimum = 0

Sum of absolute differences

$$\mathrm{SAD} = \sum_{i,j \in U} |I_1(x_L + i, y_L + j) - I_2(x_R + i, y_R + j)|$$

Disparity space image



Simple block-matching

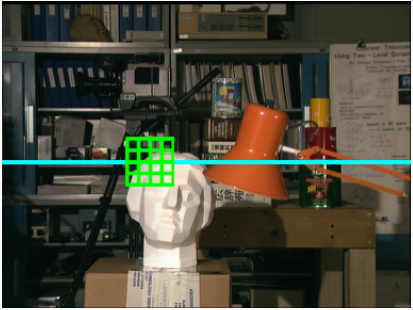
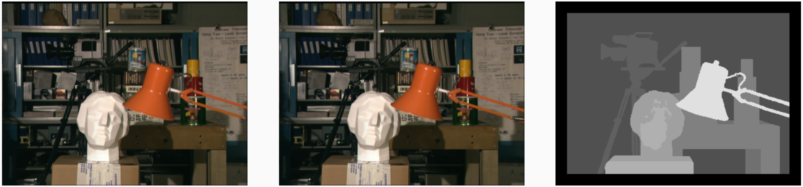


Figure 4: Illustration of block matching along a scanline.

Example for stereo image pair

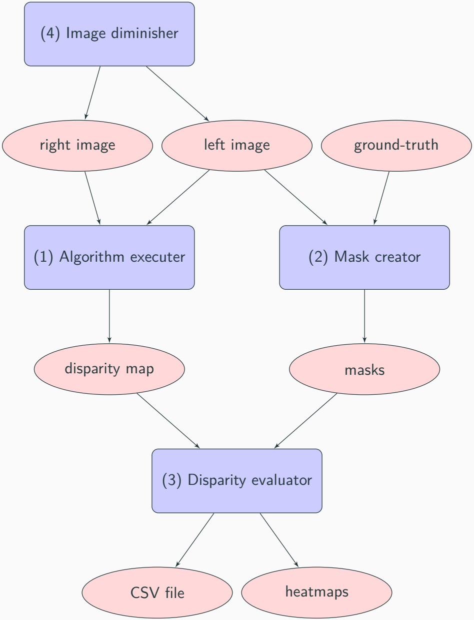


(a) left input image **(b)** right input image **(c)** ground-truth data

Figure 5: Tsukuba benchmark stereo image pair of the University of Tsukuba [1].

Implementation

Overview


(4) Image diminisher
right image
left image
ground-truth
(1) Algorithm executer
(2) Mask creator
disparity map
masks
(3) Disparity evaluator
CSV file
heatmaps

Spatiotemporal stereo matcher

Algorithm 1: CreateDisparitySpaceImage

Input: I_L, I_R, d_{max}, $wSize$

Output: C

$step \leftarrow (wSize - 1)/2$

$C \leftarrow \textsc{CreateMatrix}(\textsc{Cols}(I_L), \textsc{Rows}(I_L), d_{max})$

for $t \leftarrow 0$ ***to*** $\textsc{Images}(I_L)$ **do**

  $leftImage \leftarrow I_L(t)$

  $rightImage \leftarrow I_L(t)$

  for $y \leftarrow 0 + step$ ***to*** $\textsc{Rows}(I_L(0)) - step$ **do**

    for $x \leftarrow 0 + step$ ***to*** $\textsc{Cols}(I_L(0)) - step - d_{max}$ **do**

      for $d \leftarrow 0$ ***to*** d_{max} **do**

        $rect_L \leftarrow \textsc{Rect}\{x - step, y - step, wSize, wSize\}$

        $rect_R \leftarrow \textsc{Rect}\{x + d - step, y - step, wSize, wSize\}$

        $window_L \leftarrow leftImage(rect_L)$

        $window_R \leftarrow rightImage(rect_R)$

        $C(x, y, t, d) \leftarrow \textsc{MatchingCost}(window_L, window_R)$

return C

Spatiotemporal stereo matcher

Algorithm 2: GetDisparityMap

Input: C, t

Output: *DisparityMap*

$DisparityMap \leftarrow \text{CreateMatrix}(\text{Cols}(C), \text{Rows}(C))$
for $t \leftarrow 0$ ***to*** $\text{Frames}(C)$ **do**
    for $y \leftarrow 0$ ***to*** $\text{Rows}(C)$ **do**
        for $x \leftarrow 0$ ***to*** $\text{Cols}(C)$ **do**
            $Cost \leftarrow \frac{1}{4}C(x, y, f_0) + \frac{2}{4}C(x, y, f_1) + \frac{1}{4}C(x, y, f_2)$
            $DisparityMap(x, y) \leftarrow \text{BestMatch}(Cost)$
return *DisparityMap*

Masking modes

- Non-occluded mask
- Depth-discontinuity mask
- Textureless mask
- Saliency mask

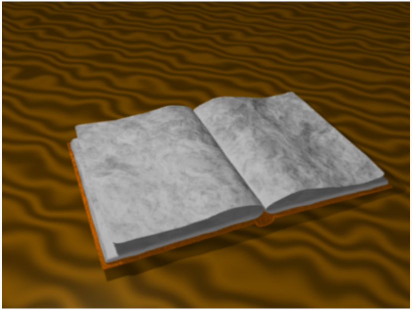


Figure 7: Frame of book sequence

Masking modes

- Non-occluded mask
- Depth-discontinuity mask
- Textureless mask
- Saliency mask

Figure 7: Ground-truth companion

Masking modes

- Non-occluded mask
- Depth-discontinuity mask
- Textureless mask
- Saliency mask

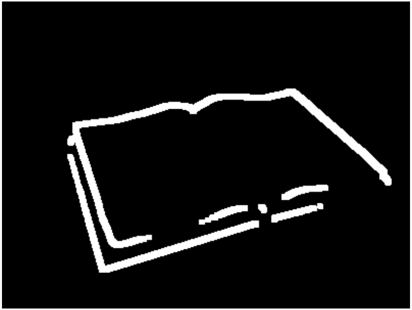


Figure 7: Depth-discontinuity at object borders

Masking modes

- Non-occluded mask
- Depth-discontinuity mask
- Textureless mask
- Saliency mask

Figure 7: Non-occluded mask

Masking modes

- Non-occluded mask
- Depth-discontinuity mask
- Textureless mask
- Saliency mask

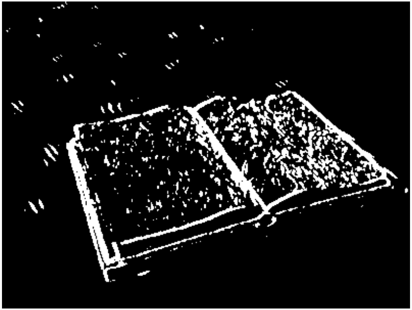


Figure 7: Textureless regions

Masking modes

- Non-occluded mask
- Depth-discontinuity mask
- Textureless mask
- Saliency mask

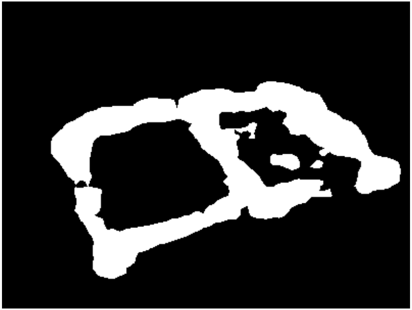


Figure 7: Salient pixels

Image diminisher

Figure 8: Flow of the image diminisher.

Image diminisher

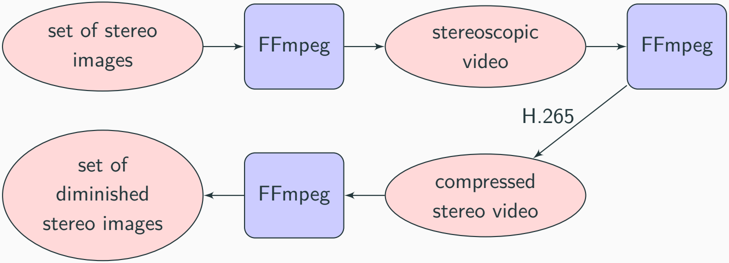


Figure 9: Flow of FFmpeg as image diminisher.

Web viewer

- Visualization of evaluation engine
- Written in Node.js
- Displaying some statistics

Demo

Evaluation

Overview

- Targeting videos (mean)
- Looking for outliers
- Masking modes
- Video compression

Datasets

- Cambridge
- SVDDD (high-resolution)

Metrics

- **Percentage of bad matching pixels**

  $\frac{1}{n}\sum_{x,y=0}(|d_a(x,y) - d_e(x,y)| > \delta_t)$

- **RMS-Error**

  $\sqrt{\frac{1}{n}\sum_{x,y=0}(d_a(x,y) - d_e(x,y))^2}$

Results

	10 CVSM	11 SNSM	12 SNTU	13 SNTW
S1	32.61%	8.72%	10.07%	9.65%
S2	25.64%	11.79%	8.76%	8.90%
S3	13.26%	6.08%	8.71%	7.29%
S4	38.96%	12.98%	11.15%	11.26%
S5	8.60%	0.93%	4.54%	2.15%
∅	23,81%	8,10%	8,66%	7,85%

Table 1: Result table for comparison of own implementation

Results

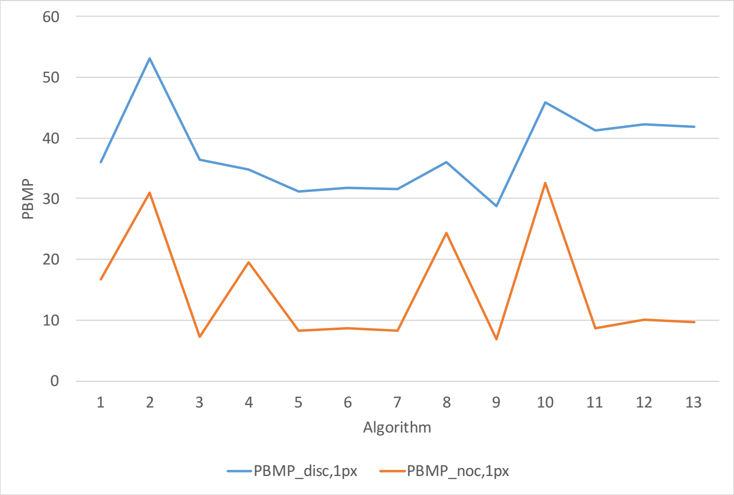


Figure 10: Depth-discontinuity mask applied on the book sequence.

Results

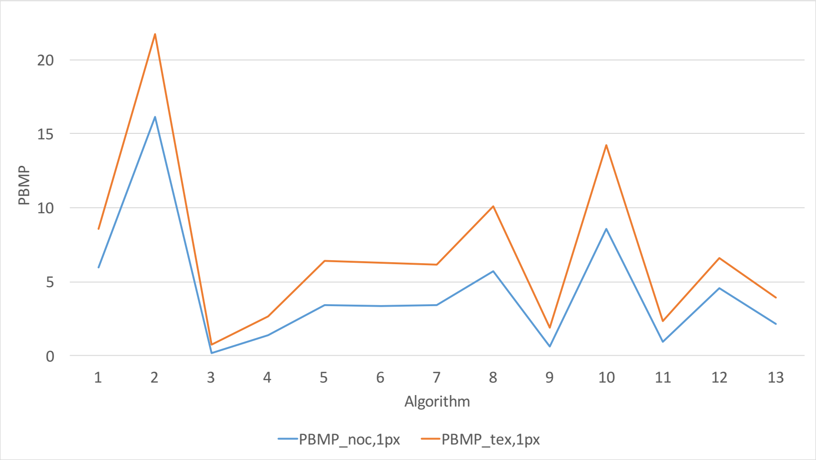


Figure 11: Chart of textureless region mask applied on the tunnel sequence.

Results

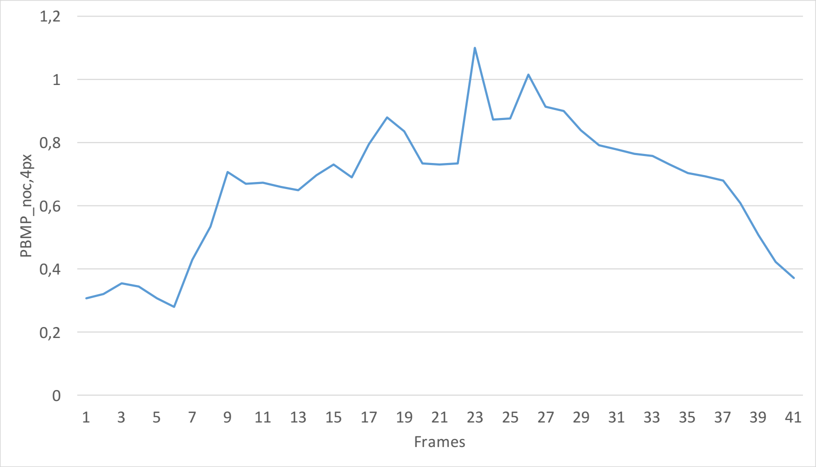
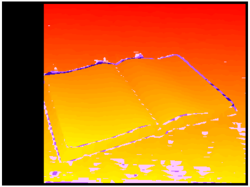
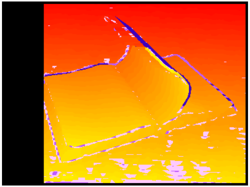


Figure 12: Chart of general outliers in a sequence.

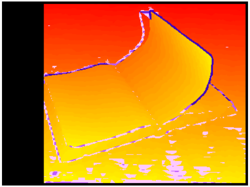
Results



(a) Frame 1



(b) Frame 23



(c) Frame 26

Figure 13: Examples for general outliers in the book sequence. The disparity maps are computed with the (3) ELAS algorithm.

Results

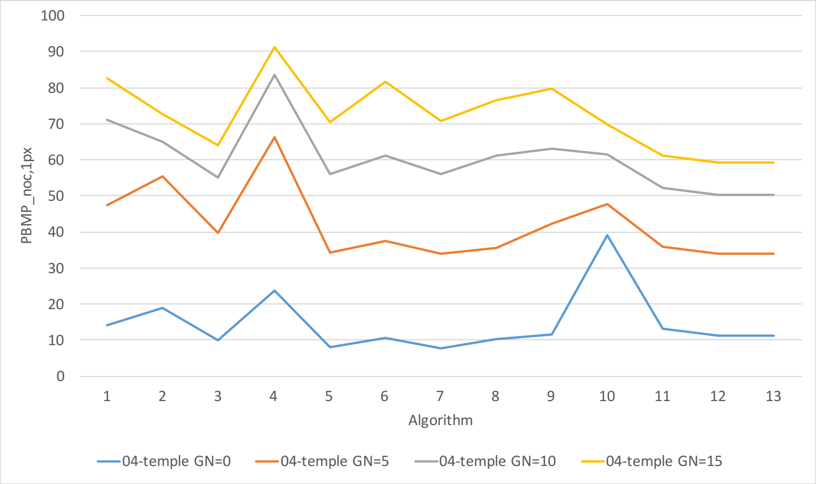
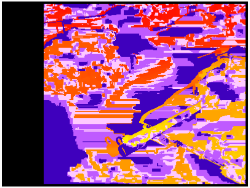
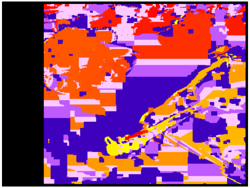


Figure 14: Chart of the impact of different σ^2 values for additive Gaussian noise on the result of disparity algorithms focusing on $P_{noc,1px}$.

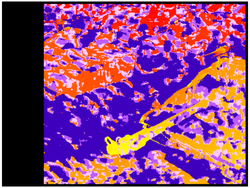
Results



(a) (3) ELAS outliers



(b) (5) MRF GC Swap outliers



(c) (13) SNSM STW

Figure 15: Example of computed disparity maps with video compression. CRF is set to 40. Frame 23 of the tanks scene.

Results

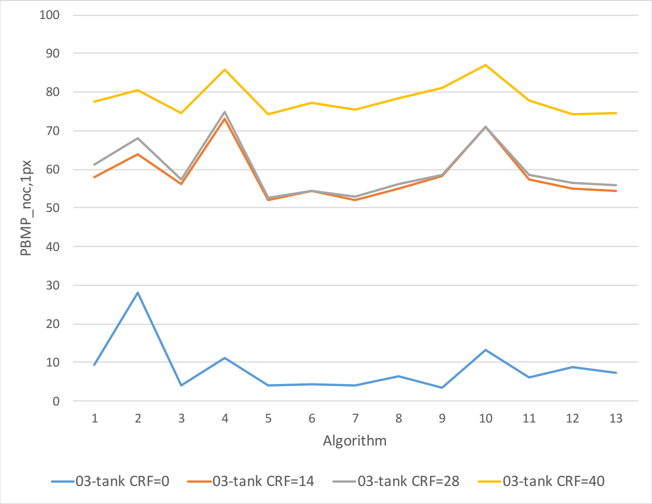


Figure 16: Chart of the impact of different CRF values for H.265 video compression on the result of disparity algorithms focusing on $\text{PBMP}_{noc,1px}$.

Results

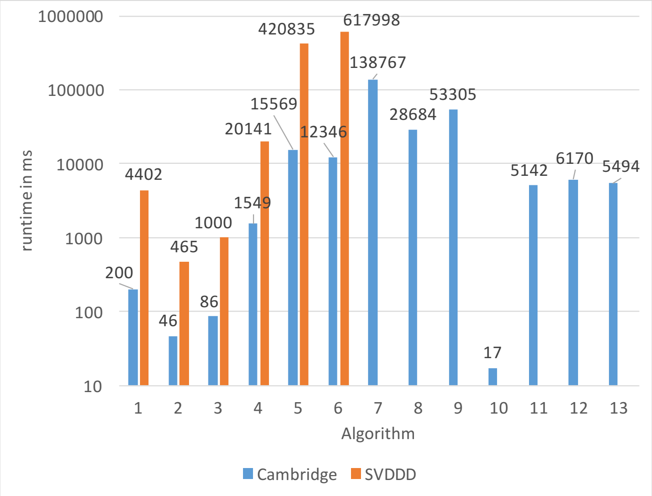
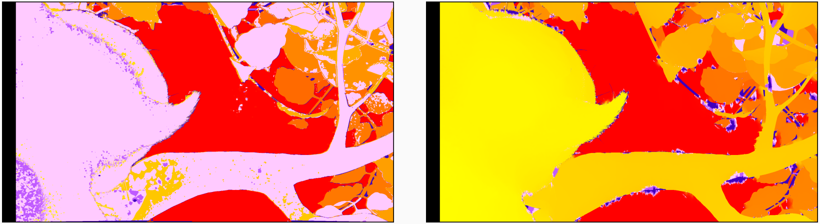


Figure 17: Comparison of the runtime of different disparity algorithms with both datasets, Cambridge and SVDDD

Results



(a) Negative disparity **(b)** Only positive disparity

Figure 18: Comparison of computed disparity maps regarding negative disparity.

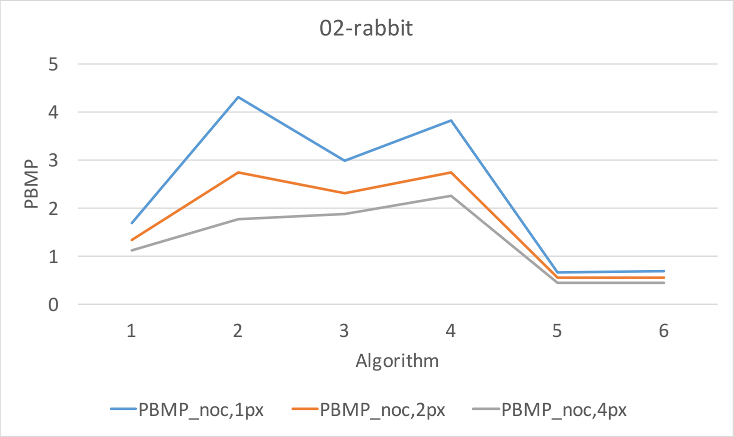


Figure 19: Performance of SVDDD rabbit scene

Results

	1	2	3	4	5	6
02-rabbit-neg	58.62%	61.51%	59.99%	60.58%	57.12%	57.13%
02-rabbit	1.68%	4.31%	2.98%	3.82%	0.65%	0.68%
03-apple	1.69%	4.10%	3.11%	3.44%	0.63%	0.65%
∅ (w/o neg)	1.69%	4.21%	3.05%	3.63%	0.64%	0.67%

Table 2: Result table for general performance of SVDDD ($\text{PBMP}_{noc,1px}$)

Conclusion

Conclusion

- surprise candidate ELAS
- camera noise model
- SVDDD dataset
- Salient mask varies a bit
- Immense runtime differences
- Possible outliers in a scene

Contributions

- PFM file reader
- Generic Disparity Interface
- Disparity Executioner
- Evaluation Engine
- Mask creator
- Image diminisher
- Python scripts
- Web result viewer
- Benchmark results
- Skeleton for stereo matcher with separated dsi
- Spatiotemporal stereo matcher

Outlook

- Motion saliency
- Enhancement of spatiotemporal matcher
- Holistic evaluation suite for modern disparity algorithm comparison

Future work

- Multi-view datasets
- High-resolution datasets
- Optical flow regarding spatiotemporal consistency
- Humans depth experience with neuronal networks

Questions?

References I

S. Martull, M. Peris, and K. Fukui.
Realistic CG stereo image dataset with ground truth disparity maps.
In *ICPR workshop TrakMark2012*, volume 111, pages 117–118, 2012.