Embedded high-resolution stereo-vision of high frame-rate and low latency through FPGA-acceleration

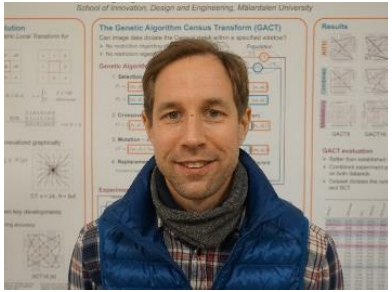
Mälardalen University

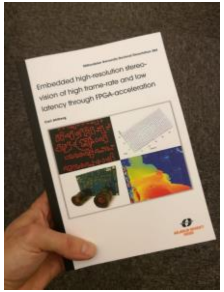
Carl Ahlberg, 2024-04-10

Outline

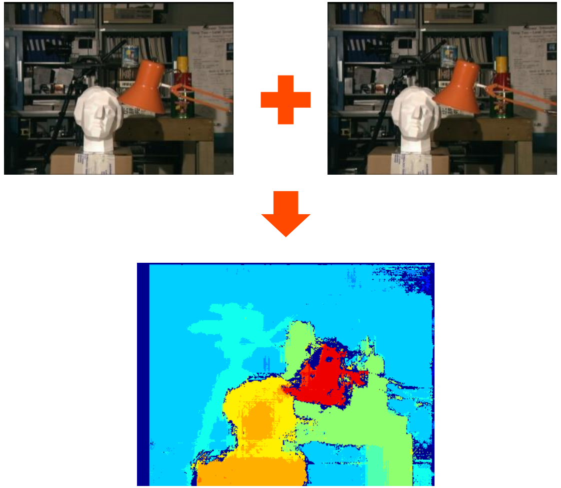
1. Presentation
2. Stereo matching
3. Census transform (possibly)
4. Segment-based matching

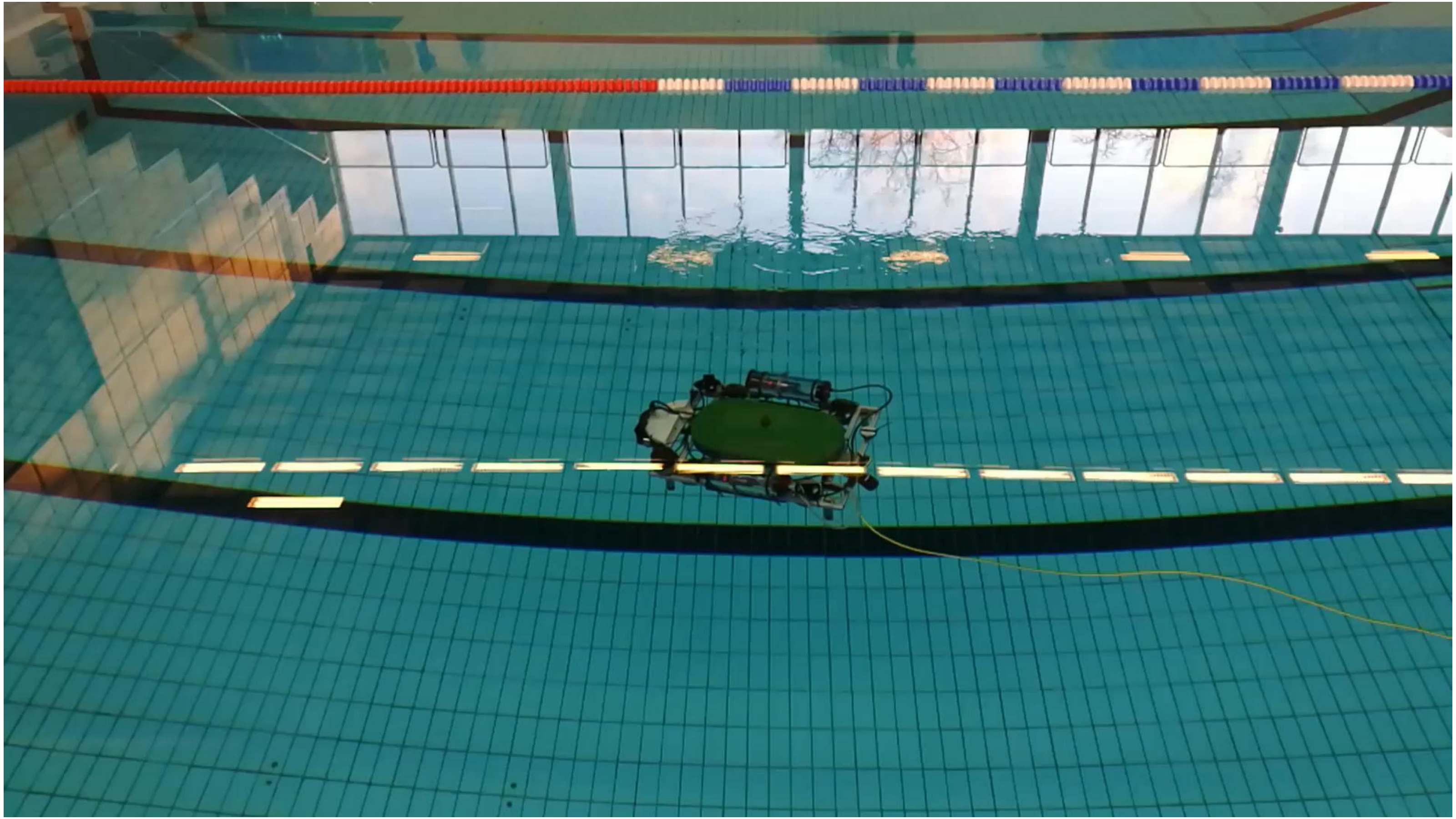
Presentation



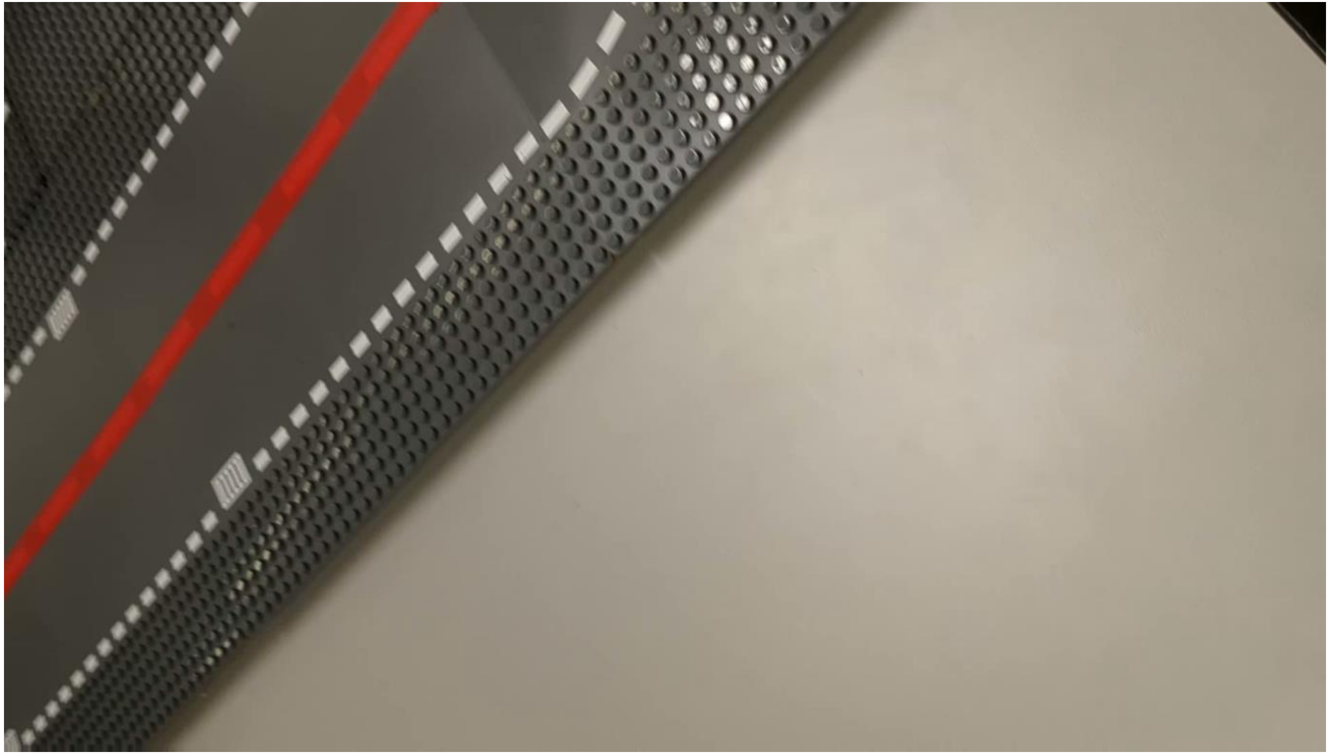


- Carl Ahlberg, Research Engineer, IFT
- PhD defense January 2020, computer science
  - *Embedded high-resolution stereo-vision of high frame-rate and low latency through FPGA-acceleration*
- Stereovision = depth from disparity
  - State-of-the-Art = AI
  - Our research = GA
- Focus – mobile robotics
  - Speed
  - Power efficiency
  - Compact
  - And… accuracy



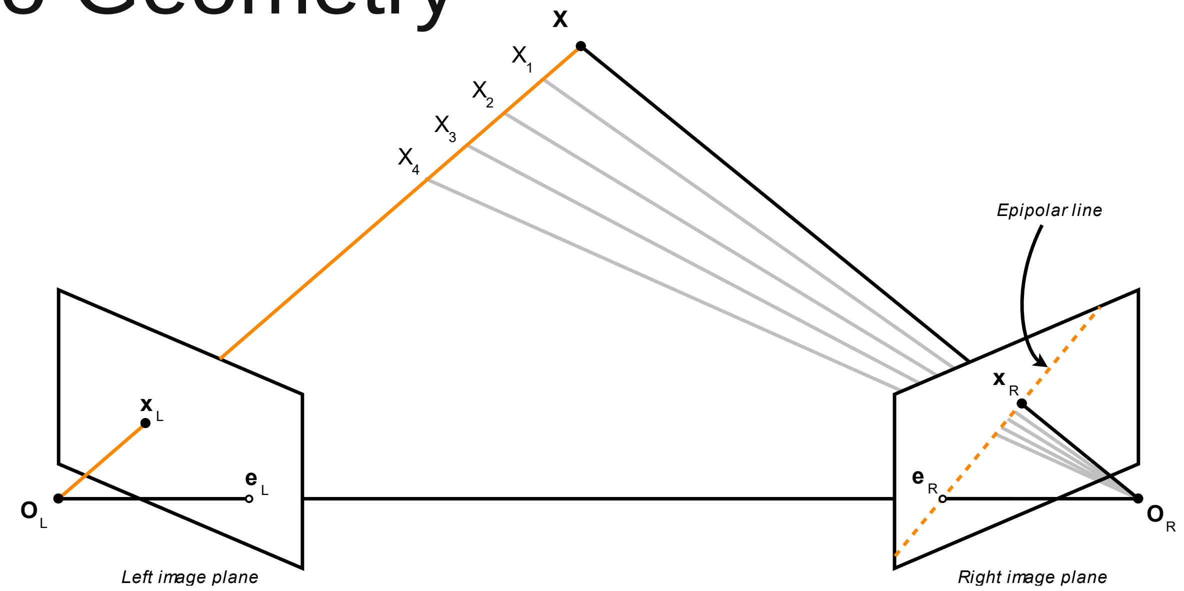


Hitachi High School (gymnasiet)



2. Stereo Matching

Stereo Geometry



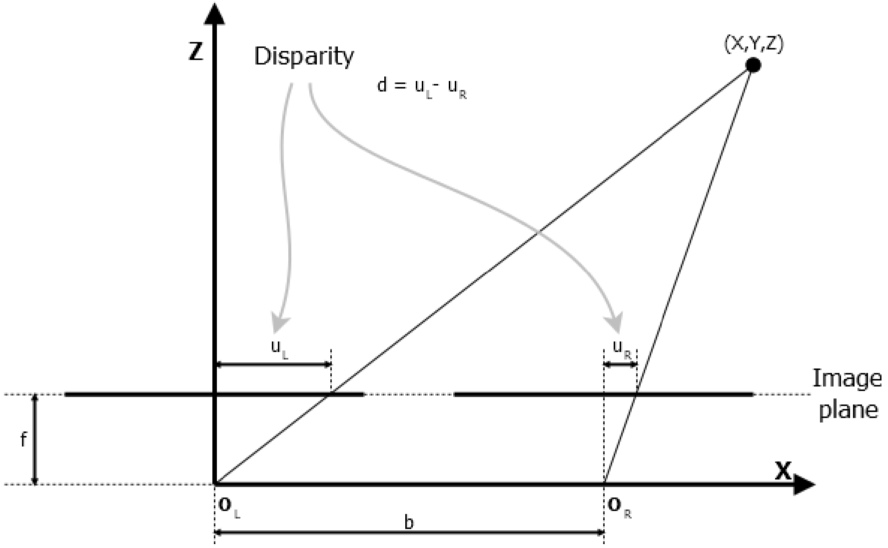
- *How to determine depth?*
- *How to find corresponding points?*

Stereo Geometry – Depth

- Assumption: rectified images
  - Corresponding points along the same line
- Disparity = Horizontal displacement
- Depth (Z) depends on disparity and camera parameters
  - Large disparity -> small depth
  - Disparity map

How to find corresponding points?

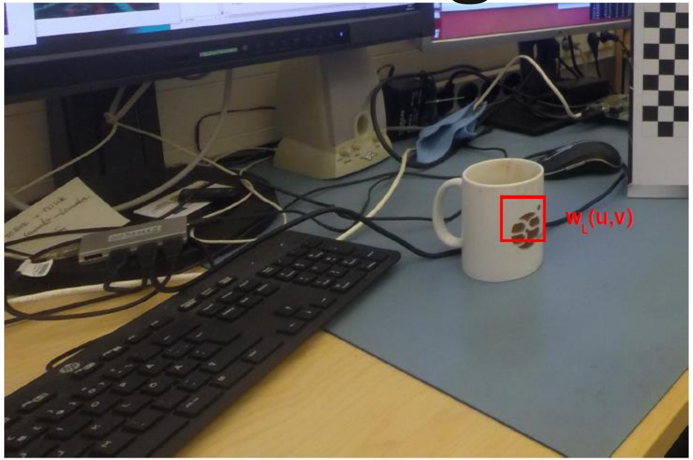
- Along X-axis: 1D search
- Limit disparity range: [0, *d_max*]

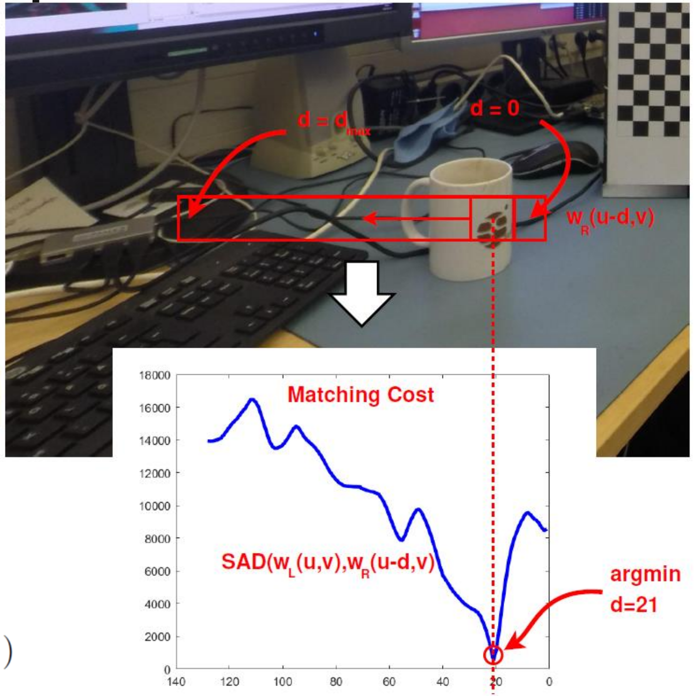


$$\frac{b}{Z} = \frac{b - u_L + u_R}{Z - f}$$

$$Z = \frac{bf}{u_L - u_R} = \frac{bf}{d}$$

Stereo Matching Example





for all u **do**
  for all v **do**
    for $i = 0 : d_{max}$ **do**
      $cost[i] = \mathrm{SAD}(w_l(u,v), w_r(u-d,v))$
    end for
    $d(u,v) = \mathrm{argmin}_i\, cost[i]$
  end for
end for

Computationally expensive!

Stereo matching (for FPGAs)

- Sparse – **Dense**
- A taxonomy of dense two-frame stereo correspondence algorithms [1]
  - **Local** – global

1. Matching cost computation (SAD, AD-GR, **Census Transform**)
2. Cost **aggregation** (Support windows, ADSW, GIF, SGM)
3. Disparity computation (WTA)
4. Disparity refinement (LRC, sub-pixel interpolation, SPF, Median filter, WMF, gap-filling)

[1] Daniel Scharstein and Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Int. J. Comput. Vision, 47(1-3):7–42, April 2002.

Part 3

Part 4

3. The Census Transform

Census Transform (CT)

"Non-parametric Local Transform for Computing Visual Correspondence" [1]

1. $$\xi(p, p') = \begin{cases} 0 & p \leq p' \\ 1 & p > p' \end{cases}$$

127	127	129
126	128	129
127	131	*A*

1	1	0
1		0
1	0	*a*

2. $$C(p) = \bigotimes_{p' \in N(p)} \xi(p, p')$$

$$[1, 1, 0, 1, 0, 1, 0, a] \qquad a = \begin{cases} 1 & A < 128 \\ 0 & otherwise \end{cases}$$

3. $$d(p_1, p_2) = \text{Hamming}\big(C(p_1), C(p_2)\big)$$

$$\begin{array}{rr} & [1,1,0,1,0,1,0,0] \\ \text{XOR} & [1,1,1,1,1,1,0,0] \\ \hline & [0,0,1,0,1,0,0,0] \end{array} \implies d = 2$$

[1] R. Zabih and J. Woodfill. Non-parametric Local Transforms for Computing Visual Correspondence. ECCV, 1994.

CT cont.

CT advantages

- Good performance along object boundaries
- Robust towards outliers
- Suitable for hardware implementation

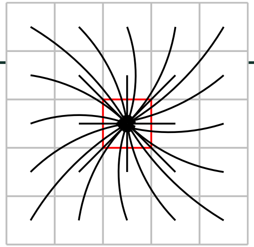
CT developments

- Sparse CTs save resources
- Non-centric comparison schema increases matching accuracy

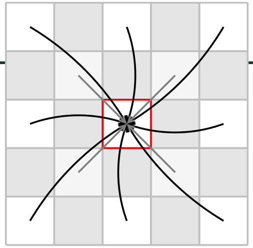
⇒Handcrafted, symmetry, edge length

⇒Can image data dictate CT schema?

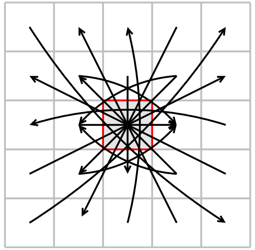
⇒Genetic Algorithm CT (GACT)



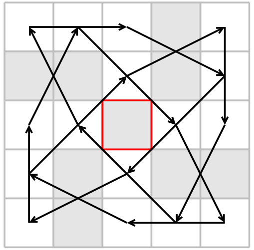
CT: n = 24, N = 5x5



Sparse CT [1]



GCT16 [2]



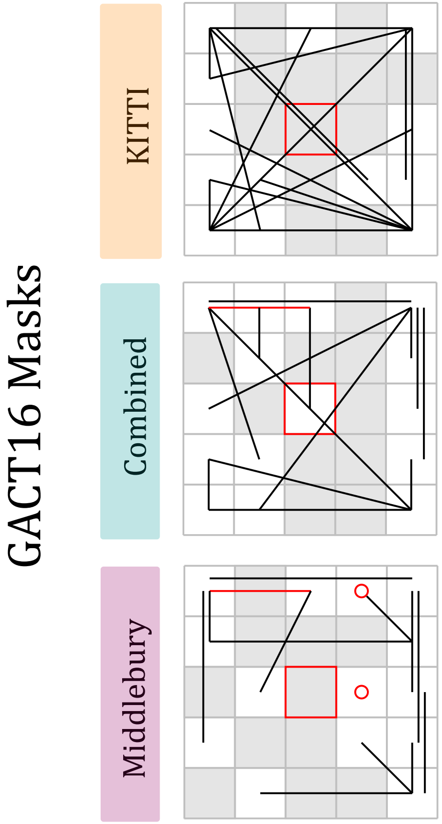
SCT16 [3]

[1] K. Ambrosch, C. Zinner, and H. Leopold. A miniature embedded stereo vision system for automotive applications. IEEEI, 2010.

[2] W. S. Fife and J. K. Archibald. Improved census transforms for resource-optimized stereo vision. TCSVT, 2013.

[3] J. Lee and H. Hong. Improved census transform for noise robust stereo matching. Optical Engineering, 2016.

GACT



KITTI

Method	Nocc	Occ	Edges
GACT24$_K$	15.45	16.88	24
GACT16$_K$	15.48	16.90	16
CT-7x9	15.80	17.22	62*
GACT2x12	16.03	17.46	24
GACT2x8	16.36	17.76	16
GACT8$_K$	16.41	17.81	8
GACT2x4	16.77	18.17	8
GCT16	17.66	19.04	16
GCT8	17.81	19.20	8
hwCSCT	18.08	19.46	14
CSCT	18.16	19.54	12
wCSCT	18.25	19.63	16
QCT	18.50	19.87	48**
SCT16-2-2	18.83	20.19	16
MeanCT	18.96	20.33	24***
SCT24-2-1	19.78	21.12	24
SCT8-2-2	20.09	21.43	8
Sparse8	20.22	21.55	8
CT	20.60	21.92	24
Sparse12	20.64	21.96	12
MCT	24.11	25.38	6

* 7 × 9 census neighborhood
** bit strings of length 48 and requires mean value calculation
*** requires mean value calculation

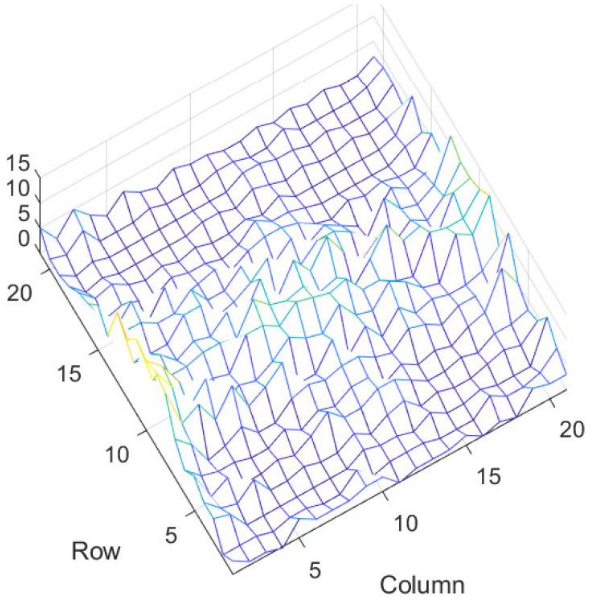
Middlebury (Nocc)

GACT2x12	**21.10**
GACT24$_M$	21.22
CT-7x9	21.27
GACT16$_M$	21.35
GACT2x8	21.37
SCT16-2-2	21.47
SCT24-2-1	21.51
GACT8$_M$	21.62
CT	21.98
Sparse12	22.18
Sparse8	22.19
GCT16	22.46
SCT8-2-2	22.49
GACT2x4	22.49
GCT8	22.62
QCT	22.65
wCSCT	23.09
hwCSCT	23.19
MCT	23.35
CSCT	23.36
MeanCT	25.00

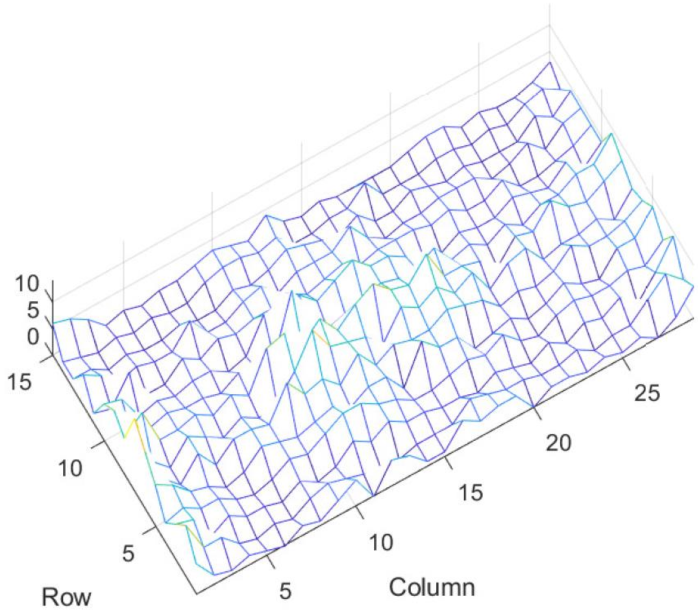
Unbounded Sparse Census Transform using Genetic Algorithm, *Carl Ahlberg, Miguel Leon, Fredrik Ekstrand and Mikael Ekström, IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa Village, Hawaii, USA, 7-11 January, 2019, pp. 1616-1625.*

GACT coordinate activation

Square 21x21



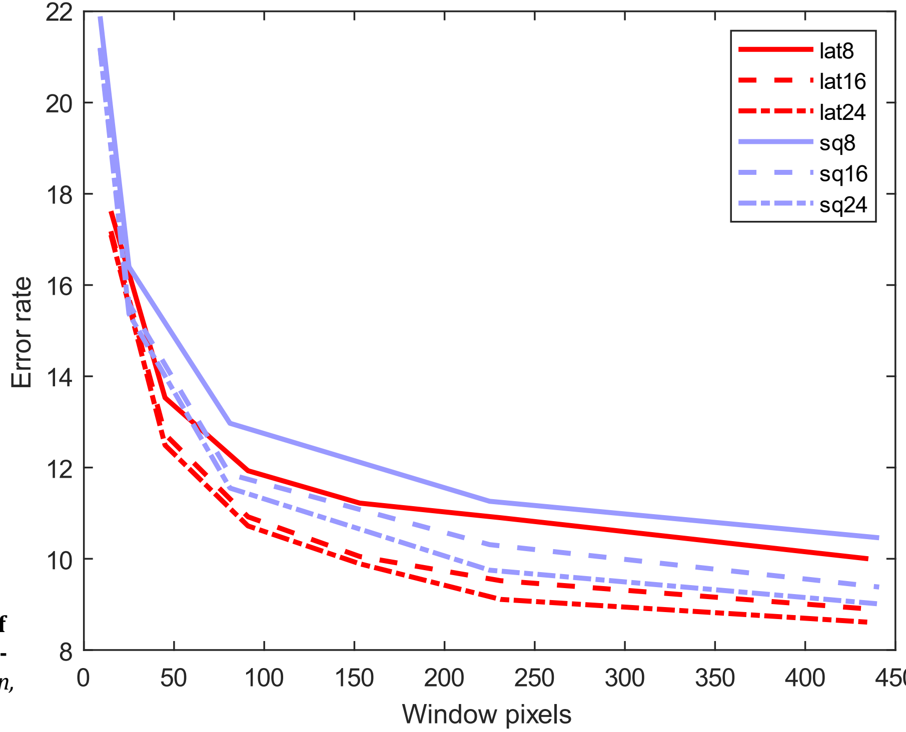
Lateral 15x29



Lateral GACT

Lateral vs square GACT

- produces better matching accuracy
- requires less resources



The Genetic Algorithm Census Transform – Evaluation of census windows of different size and level of sparseness through hardware in-the-loop training, *Carl Ahlberg, Miguel Leon, Fredrik Ekstrand and Mikael Ekström, J Real-Time Image Proc 18, 539–559 (2021), https://doi.org/10.1007/s11554-020-00993-w*

GACT vs related works

Method	Noc	Occ	Edges
GACT24 3 × 9	13.58	15.03	24[1]
GACT16 3 × 9	13.85	15.30	16[1]
GACT8 3 × 9	14.60	16.03	8[1]
GACT24 3 × 7	14.99	16.42	24[2]
GACT16 3 × 7	15.17	16.59	16[2]
GACT24 5 × 5	15.37	16.80	24
GACT16 5 × 5	15.59	17.01	16
CT 7 × 9	15.80	17.22	62[3]
GACT8 3 × 7	15.84	17.25	8[2]
GACT8 5 × 5	16.40	17.80	8
GCT16	17.66	19.04	16
GCT8	17.81	19.20	8
hwCSCT	18.08	19.46	14
CSCT	18.16	19.54	12
wCSCT	18.25	19.63	16
QCT	18.50	19.87	48[4]
SCT16-2-2	18.83	20.19	16
MeanCT	18.96	20.33	24[5]
ACT	19.43	20.77	24[6]
SCT24-2-1	19.78	21.12	24
SCT8-2-2	20.09	21.43	8
Sparse8	20.22	21.55	8
CT	20.60	21.92	24
Sparse12	20.64	21.96	12
RCT	23.20	24.48	8
MCT	24.11	25.38	6
GACT24 5 × 29	9.09	10.54	24*
GACT24 3 × 29	9.44	10.88	24**

Census Transform and Related Works

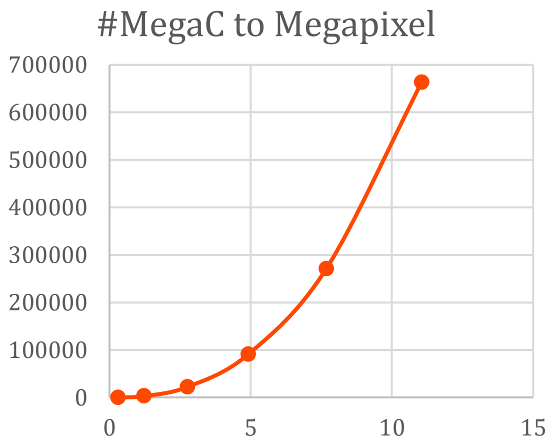
Work	Method	Image Size	D	fps	MDE/s	Platform
Gehrig 2009 [86]	ZSAD+SGM(8)	340 × 200	64	27	117.5	Virtex-4 FX140
Jin 2010 [70]	CT	640 × 480	60	230 (60)	4239 (1106)	Virtex-4 XC4LX200
Chang 2010 [81]	MCT+ADSW	352 × 288	64	42	272.5	UMC 90nm std cell (synth)
Ambrosch 2010 [32]	SAD-IGMCT	750 × 400	60	60	1080	Stratix EP1S60
Zhang 2011 [78]	MCT+VarCross	1024 × 768	64	60	3019	Stratix III EP3SL150
Banz 2011 [88]	RT+SGM(4)	640 × 480	128	30	1180	Virtex-5 LX220T-1
Greisen 2011 [71]	ZSAD	1920 × 1080	256 (64+3+3)	30	15025 (272.2)	Stratix III EP3SL340 (GPU+CPU)
Ttofis 2012 [82]	AD+segADSW	640 × 480	64	30	589	Virtex-5 LX110T
Jin M. 2012 [110]	AD-MCT+Cross+FLC	640 × 480 1024 × 768	60	507 199	9362	Virtex-6 XC6VLX240T
Perri 2012[73]	ACT+ADSW, 5 × 5 ACT+ADSW, 9 × 9	640 × 480	64 60	68 45	1337 829	Virtex- 6 XC6VLX760
Wang 2013 [93]	AD-CT+Cross+SGM(4)	1024 × 768	96	31	2400	Stratix-IV EP4SGX230
Camellini 2014 [95]	CT+SGM(8)	640 × 480	128	30	1192	Zynq Z-7020
Li 2014 [72]	CT+SAD	1920 × 1080	240	125	62208	Arria II GX EP2AGX260
Jin M. 2014 [9]	AD-MCT+Cross+FLC	640 × 480 1024 × 768	60	507 199	9362	Virtex-6 XC6VLX240T
Werner 2014 [113]	NCC+L-HRM	1920 × 1080	-	30	-	-
Ttofis 2014[83]	AD-GR+GIF+SB_WMF	1280 × 720	64	60	3538	Kintex-7 XC7K325T
Honegger 2014 [96]	CT+SGM(5)	752 × 480	32	60	693	Artix-7 XC7A100T
Schumacher 2014 [98]	CT+DD-SGM(4)	1242 × 375	160 (65)	199	14829	Virex-5
Ttofis 2015 [8]	Sparse CT+segADSW	640 × 480	64	60	1179	Kintex-7
Gherig 2015 [92]	CT+SGM(4)(sparse)	2048 × 1024	256 (64)	22	11811 (2953)	Zynq Z-7045
Pérez-Patricio 2015 [33]	AD+SBAW	1280 × 1024	16	76	1594	Cyclone II EP2C35 (DE2)
Mattoccia 2015 [100]	CT+SGM(4)	640 × 480	32	30	295	Spartan-6 LX75
Schauwecker 2015 [102]	SGM	640 × 480	112	30	1032	Xilinx Zynq
Wang 2015 [94]	AD-CT+VarCross+SGM(4)	1600 × 1200	128	42.61	10472	Stratix 5 (5SGSMD5K2)
Pérez-Patricio 2016 [79]	AD-EDGE+FUZZY	1280 × 1024	16	76	1594	Cyclone II EP2C35 (DE2)
Yang C. 2016 [84]	AD-GR+GIF	1920 × 1080	48	80	7962	Cyclone IV EP4c115
Ttofis 2016 [111]	AD-GR+GIF_WMF+SB_WMF	1280 × 720	64	60	3538	Kintex-7 XC7K325T
Cocorullo 2016 [75]	ACT(sparse)+SLBP+ADSW, (5,5)	640 × 480	60	54	995	Virtex-7 XC7VX980T
Zha 2016 [19]	Cross Trees+(LRC+Median)	1920 × 1680	60	30	5806	Kintex-7 XC7K410T (+CPU)
Li Y. 2017 [103]	AD-GR+AW-SGM(5)	1280 × 960	64	197	15492	Stratix V
Chen S. 2017 [112]	CT+sWMF	1920 × 1080	256	60	31850	Kintex-7 XC7K325T
Cambuim 2017 [104]	AD+SGM(4)	1024 × 768	128	127	12784	Cyclone IV
Michalik 2017 [77]	AD-RCT	1280 × 720	256	60	14156	Zynq XCZ7Z030
Vala 2018 [114]	DWT+AD-GR+GIF	1280 × 720	64	103	6075	Virtex-7
Rahnama 2018 [12]	CT+ELAS	-	-	10.5	-	Zynq ZC706 (CPU+FPGA)
Perri 2018 [76]	ACT(sparse)+ADSW, (3,3) ACT(sparse)+ADSW, (5,5)	640 × 480	32	81 101	796 992	Zynq XC7Z020 Zynq XC7Z045
Miyama 2019 [15]	AD+SP	1280 × 720	192	180	31850	Arria 10
Li Y. 2019 [85]	MCT+segADSW	1920 × 1080	64	45	11944	Stratix IV
Rahnama 2018 [13]	CT+R^{3}SGM	1242 × 375	128	72	4292	Zynq ZC706
Rahnama 2019 [14]	CT+R^{3}SGM+ELAS	1242 × 375	-	52	-	Zynq ZCU104 (CPU+FPGA)
Zhang X. 2019 [118]	CT+NIPM+sWMF	1920 × 1080	128	60	15925	Kintex-7 XC7K325T
Wang J. 2019 [106]	CT+SGM(4)+down-sampling	1280 × 960	75 (50)	31 42	2857 (1905) 3871 (2581)	Zynq ZC7020 Kintex-7
Ahlberg [Paper E]	GACT24(5 × 29)	1242 × 375	256	107	12800	Zynq ZCU104

- Census Transform (CT)
- CT is common for FGPA-based stereo-systems
- Improving CT may lead to improvement for all these works.
- Improvement in terms of matching accuracy and/or resource utilization

4. Segment-based matching

Block Matching Scalability

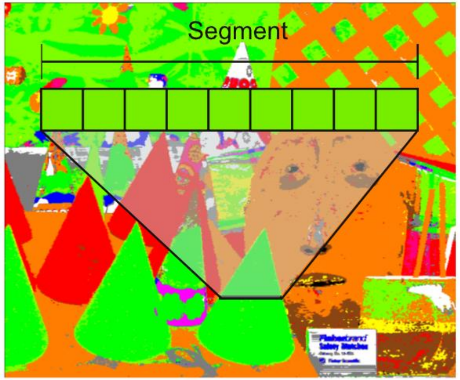
- In general stereo algorithms scale badly
- Resolution of image sensors has increased
- Affects support regions and disparity range



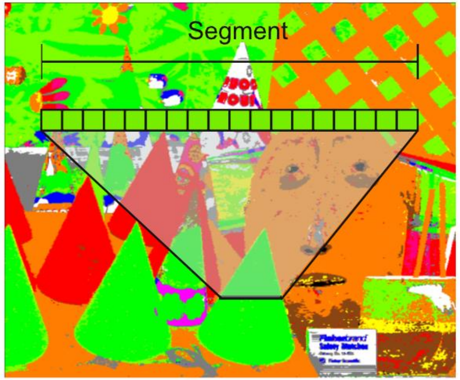
The order of calculations required to cover the same 3D-volume for increasing resolutions

Width	Height	Megapixel	#disp	Win size	#Calc(win)	#Calc(pixel)	#MegaC(img)
640	480	0.3072	16	5	25	400	123
1280	960	1.2288	32	9	81	2592	3185
1920	1440	2.7648	48	13	169	8112	22428
2560	1920	4.9152	64	17	289	18496	90912
3200	2400	7.68	80	21	441	35280	270950
3840	2880	11.0592	96	25	625	60000	663552

Segment-based matching



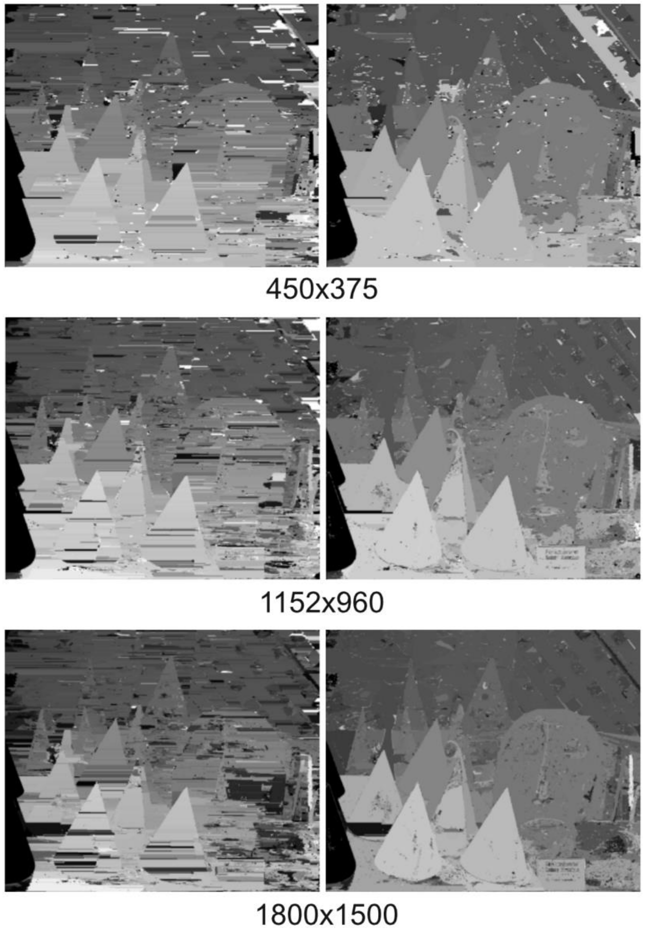
Low Res



Hi Res

- Segments are consistent over resolutions
- #segments << #pixels
- Improve estimates in textureless regions
- Provide both disparty and segment information

Initial (FGPA) Median (CPU)



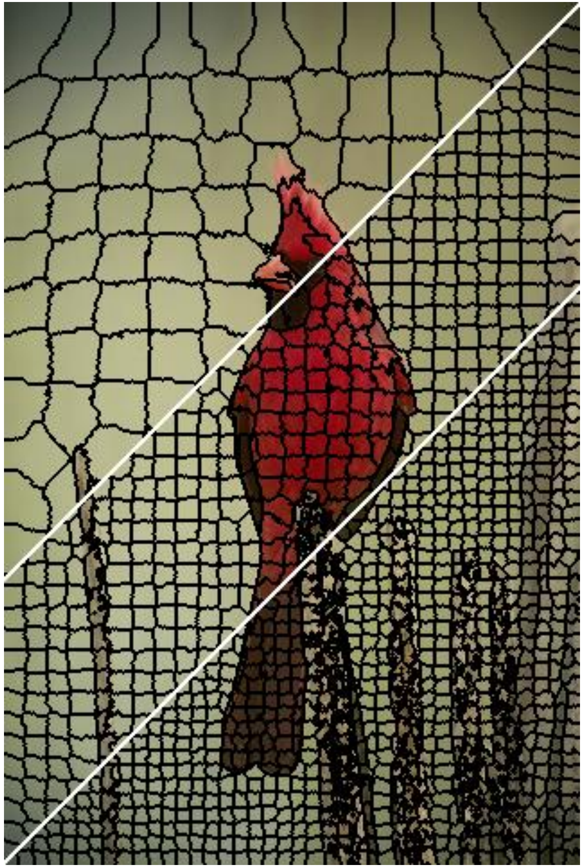
Towards an Embedded Real-Time High Resolution Vision System

Fredrik Ekstrand, **Carl Ahlberg**, Mikael Ekström, Giacomo Spampinato, *10th International Symposium on Visual Computing (ISVC 2014),* Las Vegas, Nevada, USA, 7-9 December, 2014.

- Pipelined FPGA-implementation
- 70 MHz => 2MP@35 fps
- **1800 x 1500, 240 disp, ~26 fps**

Superpixels

- A superpixel is a cluster of pixels based on
  - Similarity (color)
  - Proximity
- A preprocessing step for
  - Segmentation (different objects in the image)
  - Stereo matchning
- On FPGA
  - *FP-SLIC: A Fully-Pipelined FPGA Implementation of Superpixel Image Segmentation*, EUROMICRO DSD (Digital System Design) 2022



Mälardalen University

Thank you!

R. C. FEORD, M. E. SUMNER, S. PUSDEKAR, L. KALRA, P. T. GONZALEZ-BELLIDO, TREVOR J. WARDILL,
Cuttlefish use stereopsis to strike at prey*,*
SCIENCE ADVANCES 08 JAN 2020 : EAAY6036

https://advances.sciencemag.org/content/6/2/eaay6036