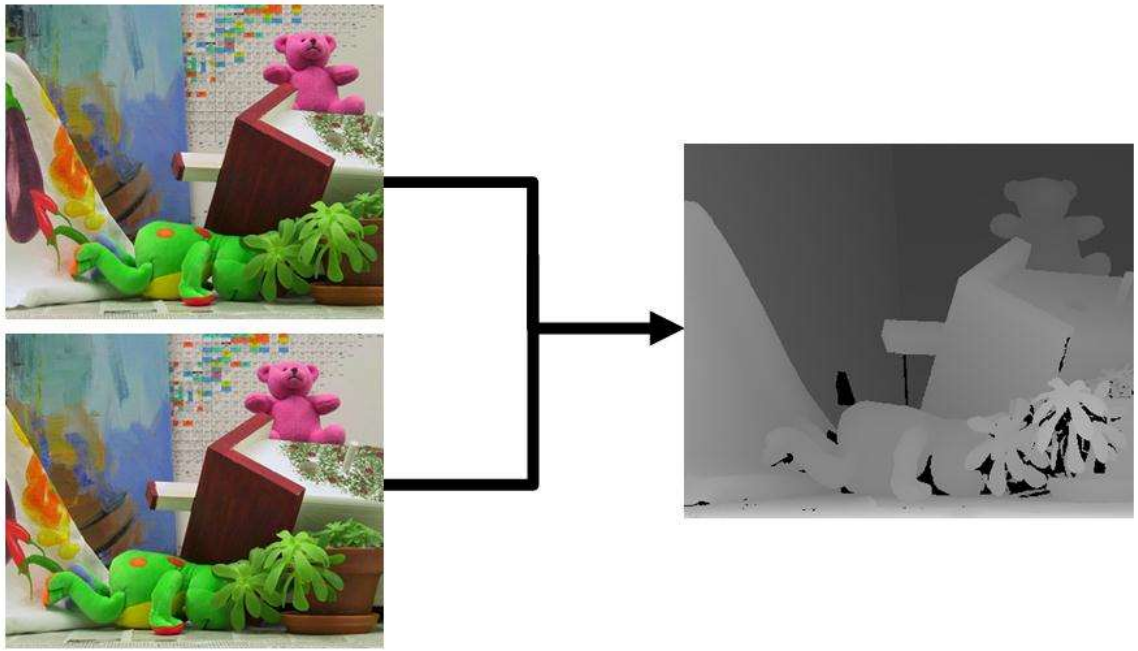
A FAST AND HIGH QUALITY STEREO MATCHING ALGORITHM ON FPGA

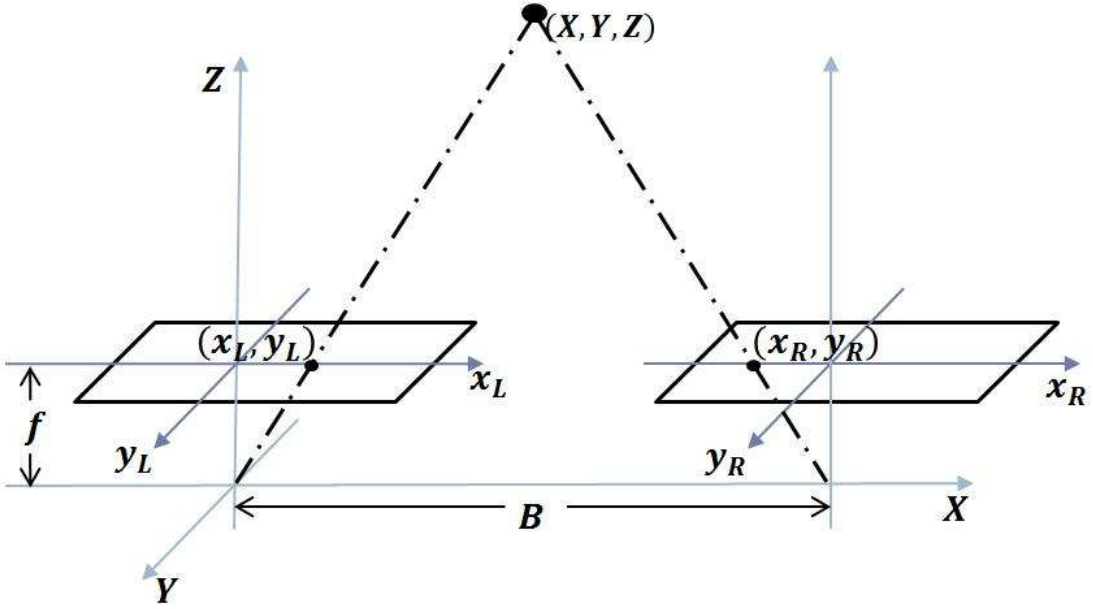
FPL2012
August 29-31, 2012
Oslo, Norway

Minxi Jin and Tsutomu Maruyama
University of Tsukuba



Background

- The aim of stereo vision systems is to reconstruct the 3-D geometry of a scene from images of two separate cameras. The major objective of stereo matching is to find the matching points in both left and right image in real time.



$$(x_L, y_L) = \left(f\frac{X}{Z}, f\frac{Y}{X}\right)$$

$$(x_R, y_R) = \left(f\frac{X-B}{Z}, f\frac{Y}{X}\right)$$

$$\Downarrow$$

$$d = x_L - x_R = f\frac{B}{Z}$$

$$\Downarrow$$

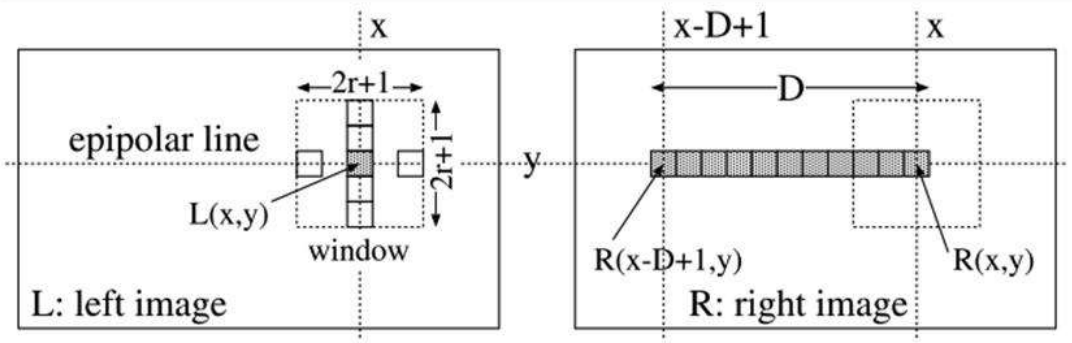
$$Z = f\frac{B}{d}$$

Earlier Studies

- Correlation-based Algorithm (ex. SSD, SAD, Census)
  - Simple Algorithm & High Speed & High Error Rate
- Dynamic Programming (DP)
  - DP on the 1D scan line structure suffers from the "streaking" problem.
- Belief Propagation (BP) & DP-Tree
  - The error rates are lower than the correlation-based algorithm, because the pixels are compared on 2D. But the two-dimensional scanning of the image limits their processing speed.
- Improvement in this study
  - We aim to construct a stereo vision system with low error rate (comparable with the top-level software algorithms) and high processing speed (which can compete with the fast stereo vision systems).

Our Method: Matching Cost-1 (Mini-Census & AD)

- $C_{MC+AD}(x,y,d) = C_{MC}(x,y,d) + \beta \times C_{AD}(x,y,d)$
  - $C_{MC}(x,y,d) = H_{amming}D_{istance}(MC(L,x,y), MC(R,x-d,y)$
    - $MC(L,x,y) = \{L(x,y-2) \geq L(x,y), L(x,y-1) \geq L(x,y), L(x-2,y) \geq L(x,y), L(x+2,y) \geq L(x,y), L(x,y+1) \geq L(x,y), L(x,y+2) \geq L(x,y)\}$
  - $C_{AD}(x,y,d) = 1 - \exp(-\frac{|L(x,y)-R(x-d,y)|}{\lambda})$



The calculation method of $C_{MC}(x,y,d)$

Matching Cost-2 (Cost Aggregation)

- First, the aggregate cost along the x axis is calculated as

$$CA_x(x,y,d) = \sum_{dx=-4}^{+3} C_{MC+AD}(x+dx,y,d)$$

$$CA_x(x,y-1,d) = \sum_{dx=-4}^{+4} C_{MC+AD}(x+dx,y-1,d)$$

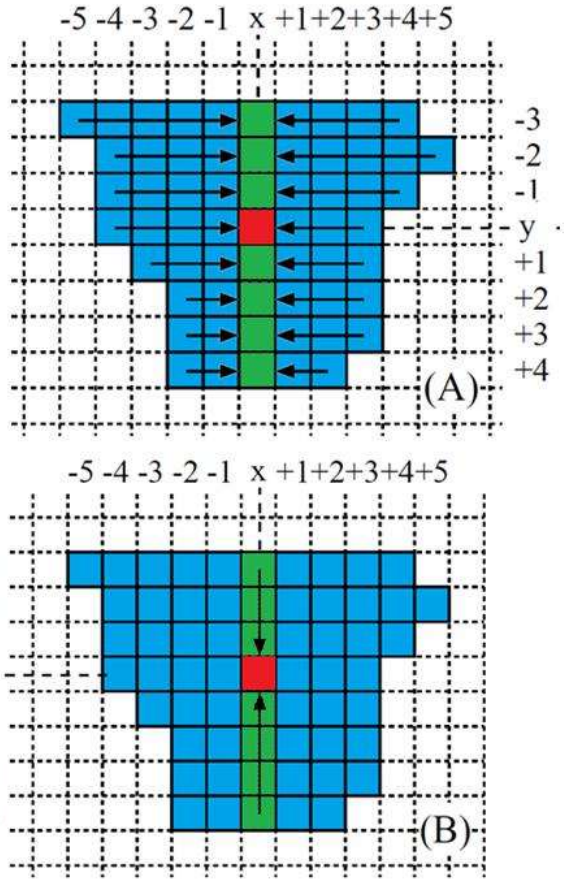
$$\vdots$$

- Then, the aggregated cost at (x,y) is given by

$$CA(x,y,d) = \sum_{dx=-3}^{+4} CA_x(x,y+dy,d)$$

- d which minimizes CA(x,y,d) is chosen as the disparity at L(x,y)

$$D_{map}(L,x,y) = \min_k CA(x,y,k)$$

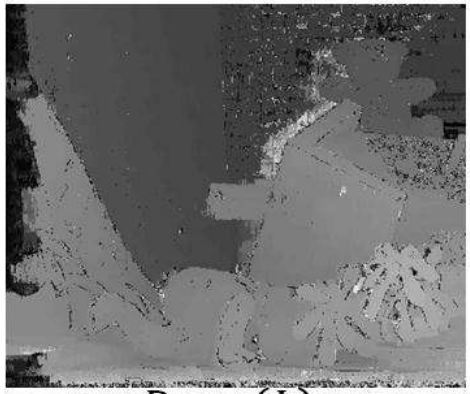


The colors of the blue and green pixels are similar to the color of the red pixel

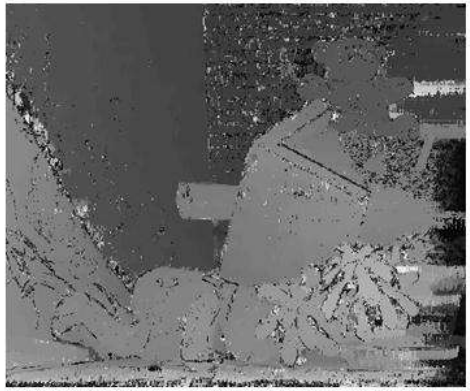
Detection of the GCPs (Ground Control Pixels)

- We want to find pixels which have high reliability from $D_{map}(L, x, y)$ and $D_{map}(R, x, y)$.

$$D_{map}(R, x, y) = k \;\&\; k - 1 \leq D_{map}(L, x + k, y) \leq k + 1$$

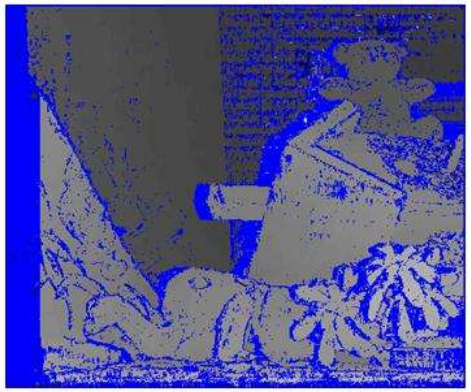


$D_{map}(L)$

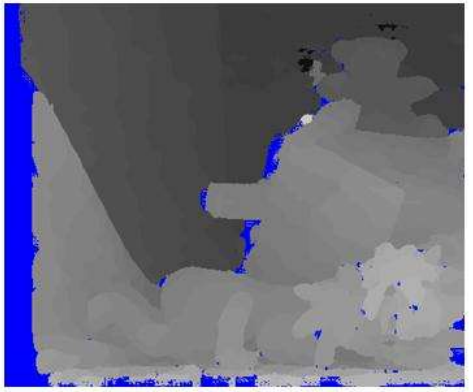


$D_{map}(R)$

After detection of the GCPs



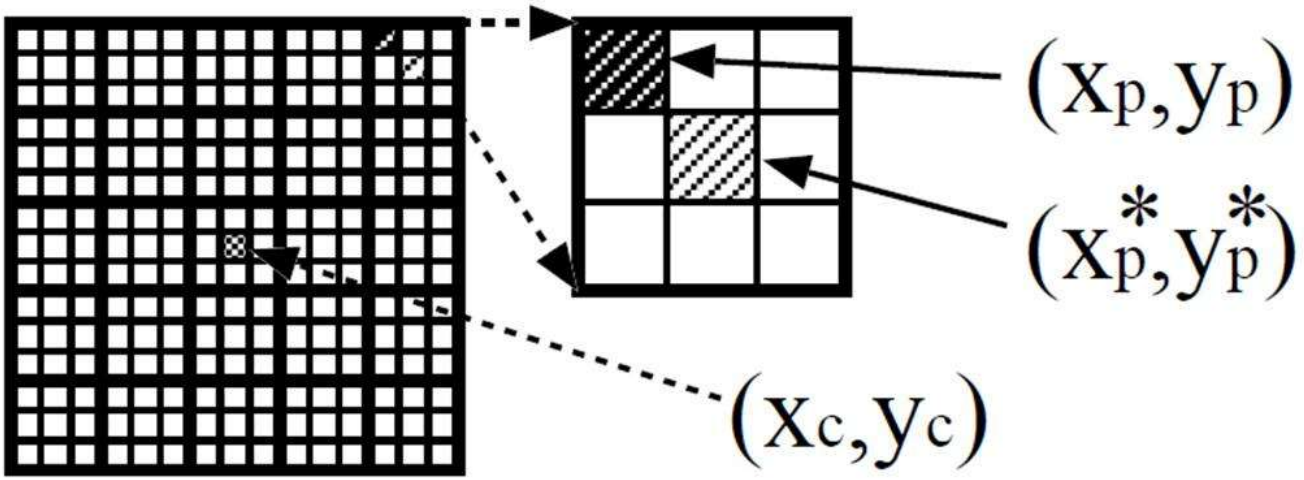
After FLC algorithm



The blue pixels are the pixels with uncertain disparity

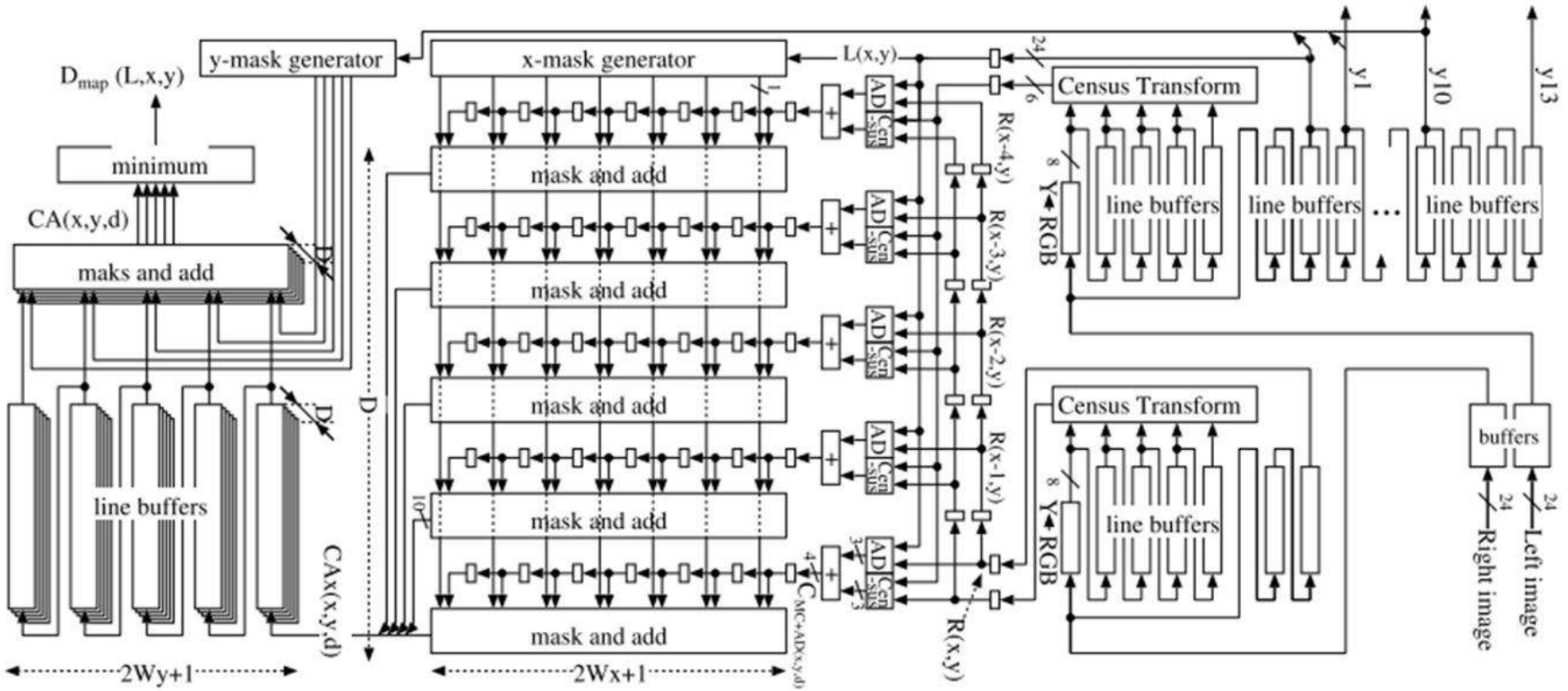
FLC algorithm (Fast Locally Consistent)

- We want to improve the accuracy of the GCPs by calculating the reliability of the GCPs.
  - $R(x_c, y_c, x_p, y_p, d) = f_{dist}(x_c, y_c, x_p^*, y_p^*) \times f_{color}(L(x_c, y_c), L(x_p^*, y_p^*)) \times f_{color}(L(x_p, y_p), R(x_p - d, y_p))$
    - $f_{dist}(x_c, y_c, x_p, y_p) = 1 - \exp(-\frac{\sqrt{(x_p - x_c)^2 + (y_p - y_c)^2}}{\lambda_d})$
    - $f_{color}(p, q) = 1 - \exp(-\frac{|p-q|}{\lambda_c})$



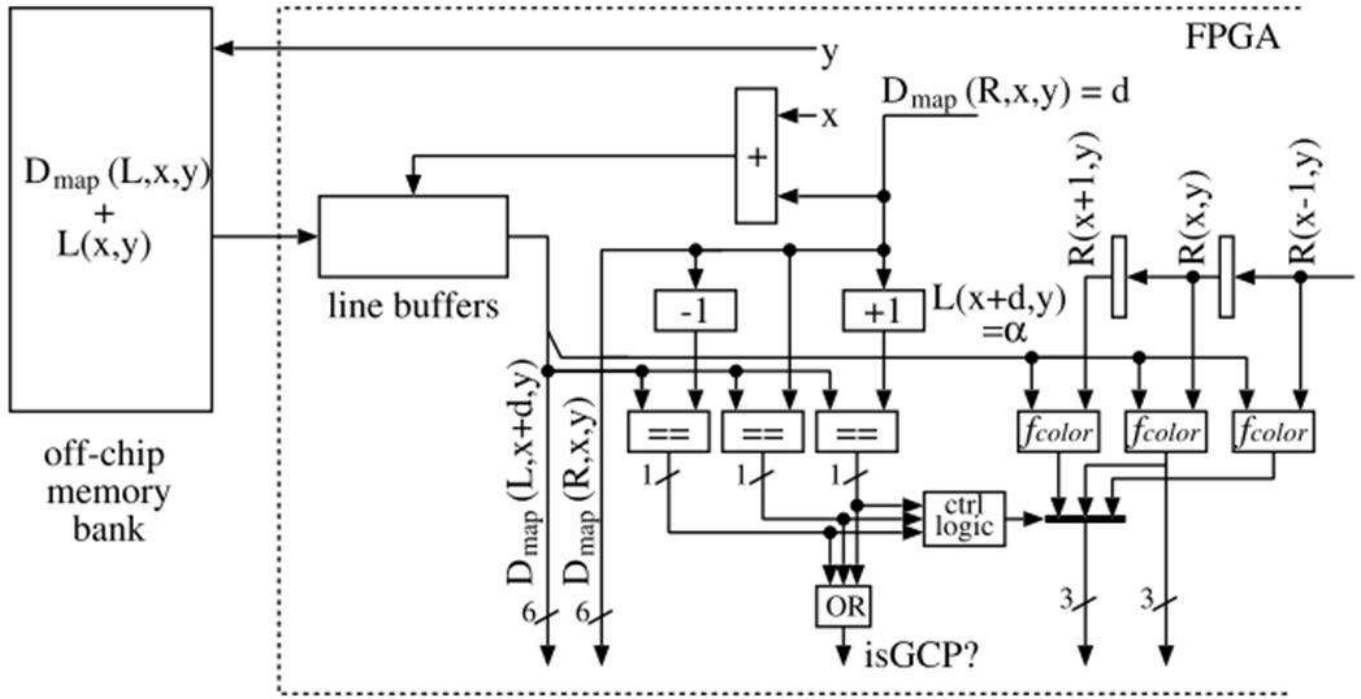
Hardware Design

- Cost Aggregation Unit



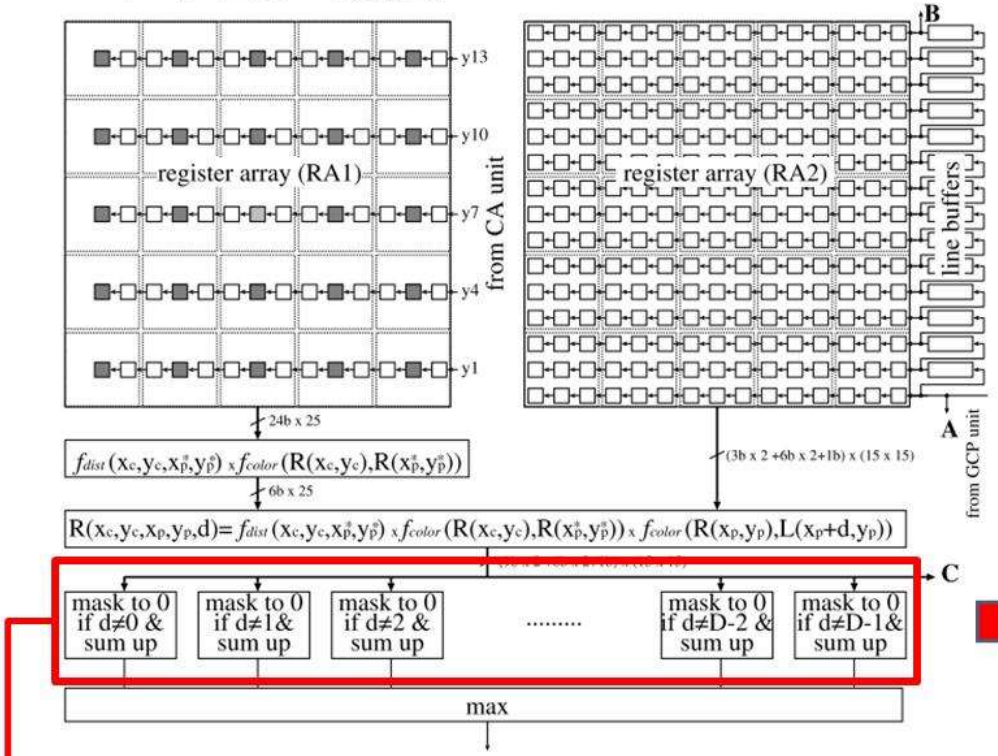
Hardware Design

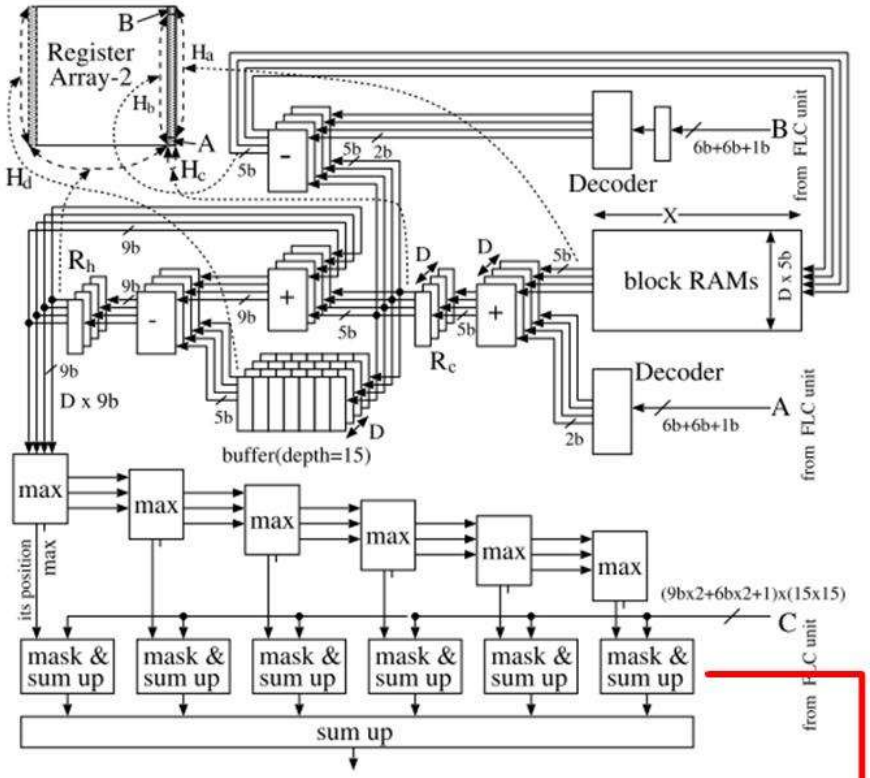
- GCP Unit



Hardware Design

- FLC Unit

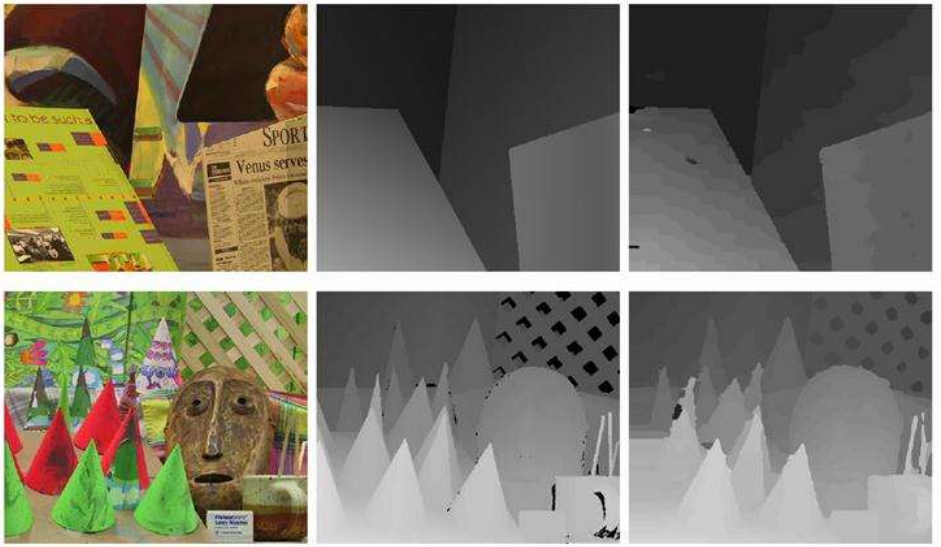




26640 adders are necessary when D=60

By using a histogram which counts the number of the disparities and summing up for only the six disparities from the top, only 2904 adders are necessary.

Experimental Results



Left: Original image Middle: True disparity map Right: Disparity map by our system

Experimental Results

algorithm	Tsukuba			Venus			Teddy			Cones		
	n.o.	all	disc	n.o.	all	disc	n.o.	all	disc	n.o.	all	disc
ADCensus[9]	1.07	1.48	5.73	0.09	0.25	1.15	4.10	6.22	10.9	2.42	7.25	6.95
AdaptingBP[10]	1.11	1.37	5.79	0.10	0.21	1.44	4.22	7.06	11.8	2.48	7.92	7.32
GeoDif[11]	1.88	2.35	7.64	0.38	0.82	3.02	5.99	11.3	13.3	2.84	8.33	8.09
our method	1.38	1.84	7.36	0.30	0.48	2.09	7.41	12.7	17.5	3.44	9.19	9.90
RealTimeBP[12]	1.49	3.40	7.87	0.77	1.90	9.00	8.72	13.2	17.2	4.61	11.6	12.4
MiniCensus+CA[5]	3.84	4.34	14.2	1.20	1.68	5.62	7.17	12.6	17.4	5.41	11.0	13.9
RealTimeDP-Tree [4]	1.43	2.51	6.60	2.37	2.97	13.1	8.11	13.6	15.5	8.12	13.8	16.4
RealTimeGPU[13]	2.05	4.22	10.6	1.92	2.98	20.3	7.23	14.4	17.6	6.41	13.7	16.5
BP+MLH[14]	4.17	6.34	14.6	1.96	3.31	16.8	10.2	18.9	24.0	4.93	15.5	12.3
DP[15]	4.12	5.04	12.0	10.1	11.0	21.0	14.0	21.6	20.6	10.5	19.1	21.1
SSD+MF[15]	5.23	7.07	24.1	3.74	5.16	11.9	16.5	24.8	32.9	10.6	19.8	26.3
Census[16]	9.79	11.6	20.3	3.59	5.27	36.8	12.5	21.5	30.6	7.34	17.6	21.0

1. n.o.(non-occluded regions) are the errors are only for the non-occluded regions.
2. all (all regions) are the errors in all regions (excluding borders of the image).
3. disc (discontinuity) are the errors only for the regions near depth discontinuities.

Experimental Results

- Platform
  - Xilinx XC6VLX240T

- Circuit size and the performance

LUTs	122.9K(81%)	
Block RAMs (36Kb)	198(47.6%)	
Operational frequency	318.259MHz	
Performance	199.7fps	1024x768
	507.4fps	640x480

Summary & Future work

- The error rate obtained by our system is close to the top-level software algorithms, and its processing speed is comparable with the fastest FPGA stereo vision systems.
- We have demonstrated how accurate stereo matching is possible by not using global matching algorithms.

As the next step,

- We want to introduce a part of the global matching algorithms, and reducing the circuit size by considering the balance of the processing speed and the circuit size.