Video-Rate Stereo Vision on a Reconfigurable Hardware

Ahmad Darabiha

Department of Electrical and Computer Engineering

University of Toronto

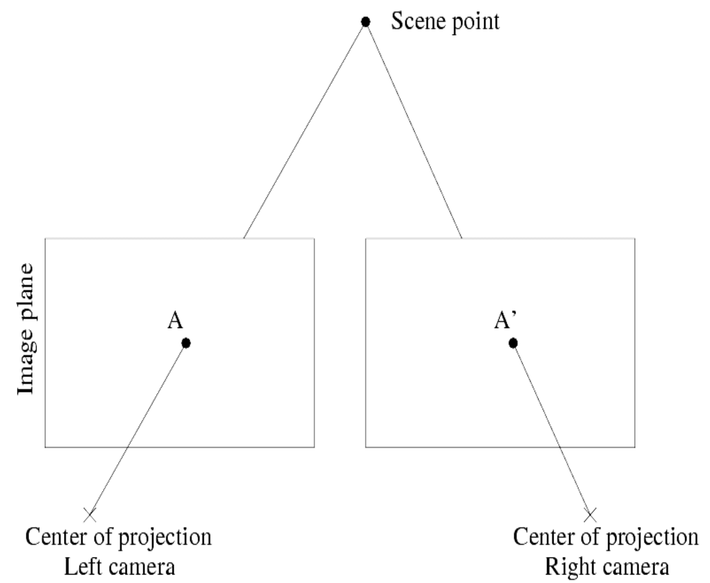
Introduction

- What is “Stereo Vision”?

“The ability of finding the *depth information* encoded within *multiple images*”

- Applications?

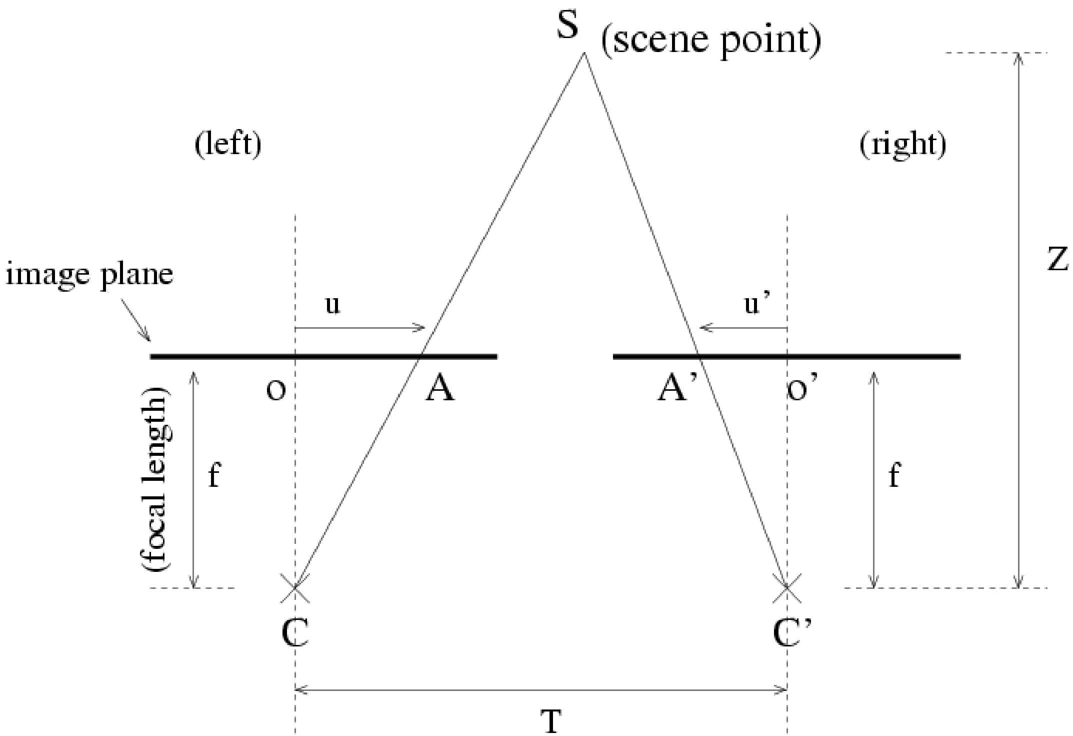
  - Robotics, Navigation
  - Security, Monitoring



Motivation

- Problem
  - Real-time vision applications → 30 frames/sec
  - Fastest software systems 5-10 seconds for each frame
- Solution
  - Hardware implementation can accelerate the performance to video rate

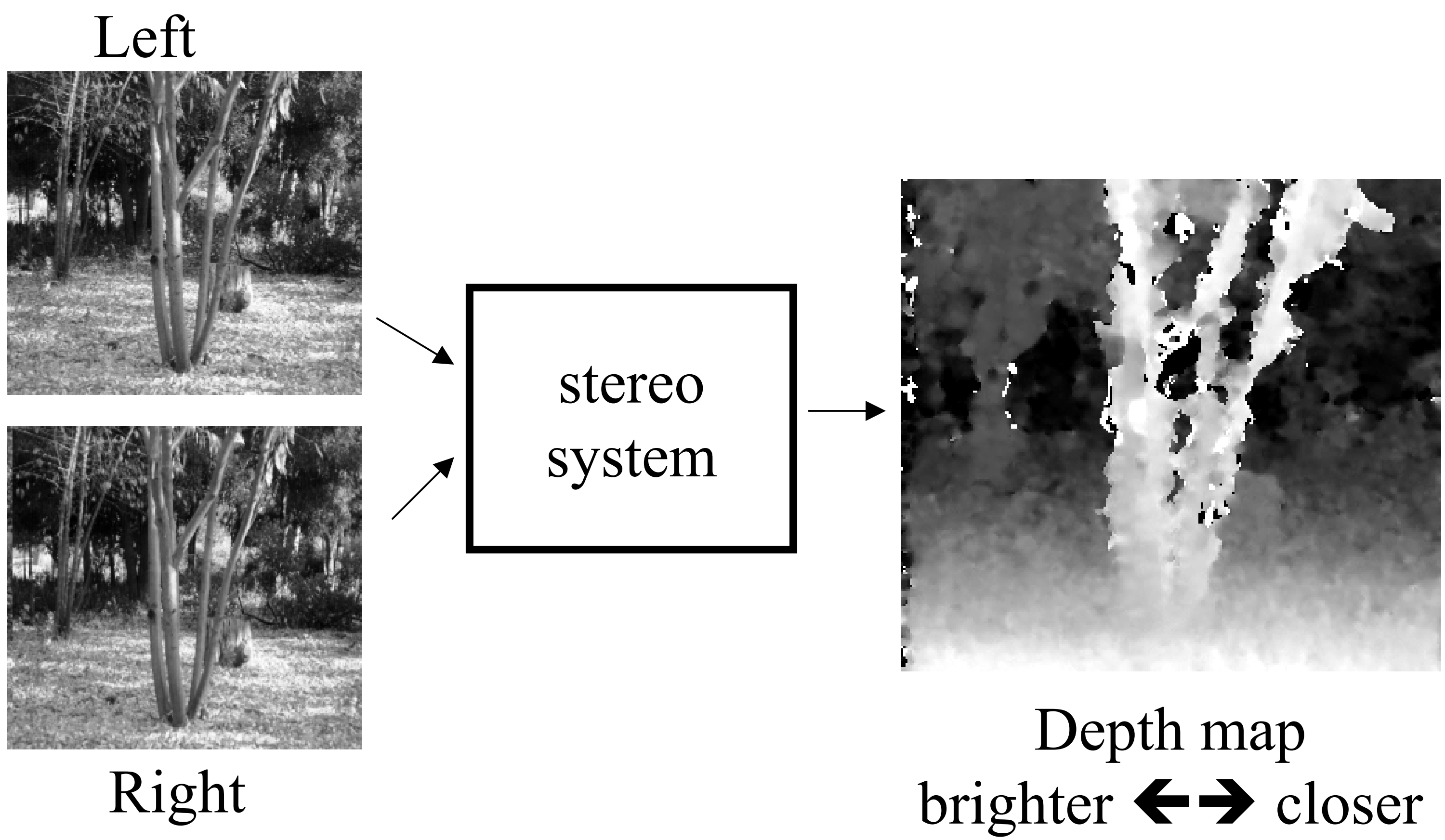
Stereo Basics



Top view

- f : focal length
- T : distance between cameras
- Disparity $d = u - u'$
- Distance $Z = f\,T/d$

Example



How to find the corresponding points?

Correspondence Problem

How to match corresponding
points between the two images?

Three methods:

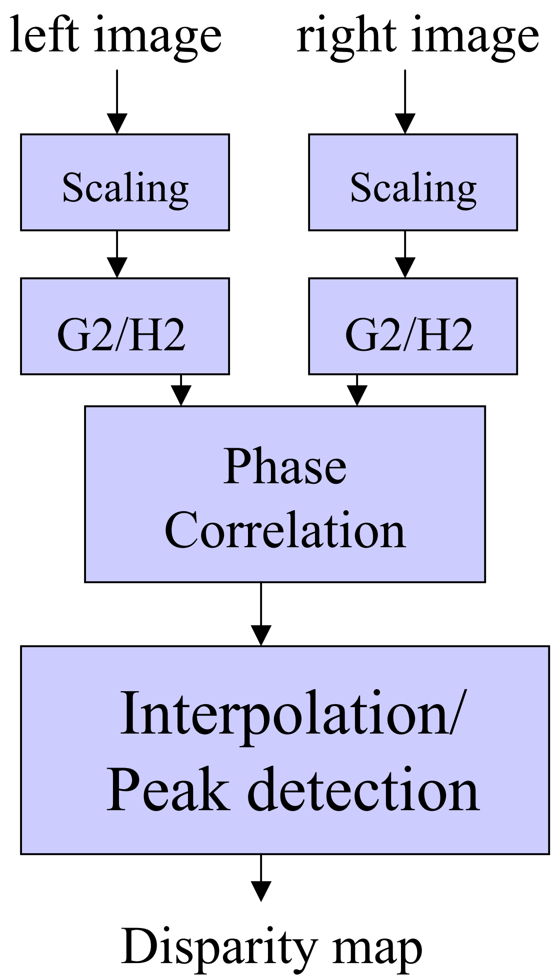
- *Intensity-based*
  - Match the pixels based on their intensity values
  - ✖ Sensitive to brightness variations
- *Feature-based*
  - Edges, corners, straight lines
  - ✖ Can not produce dense disparity maps
- *Phase-based*
  - Phase of filter outputs
  - ✓ Brightness invariant
  - ✓ Extracts more local texture

Local-Weighted Phase Correlation Algorithm

- Adopted in our system
- Phase-based
  - G2/H2 filters to extract the phase
- Multi-resolution
  - Will reduce false matches
  - Three scales: 1,2 and 4
- Multi-orientation
  - Extracts more texture
  - Directions –45, 0, 45 degrees

Local-Weighted Phase Correlation Algorithm

- Four major steps:
  1. Scaling
  2. Orientation Decomposition
  3. Phase Correlation
  4. Interpolation/ Peak-Detection



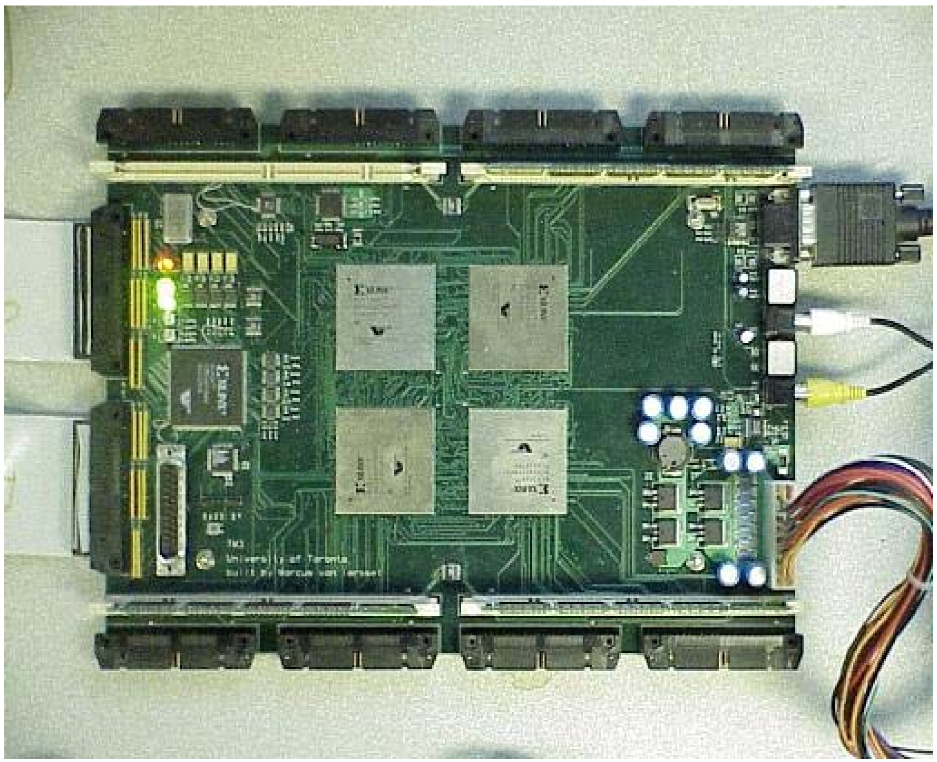
Hardware Design

Hardware: ASIC or FPGA?

- ✘ ASIC (Application Specific Integrated Circuit)
  - Expensive and long design cycle
  - Preferred in mass production

- ✓ FPGA (Field-Programmable Gate Array)
  - Less stringent design cycle
  - Less expensive
  - Can change the circuit "on the fly"

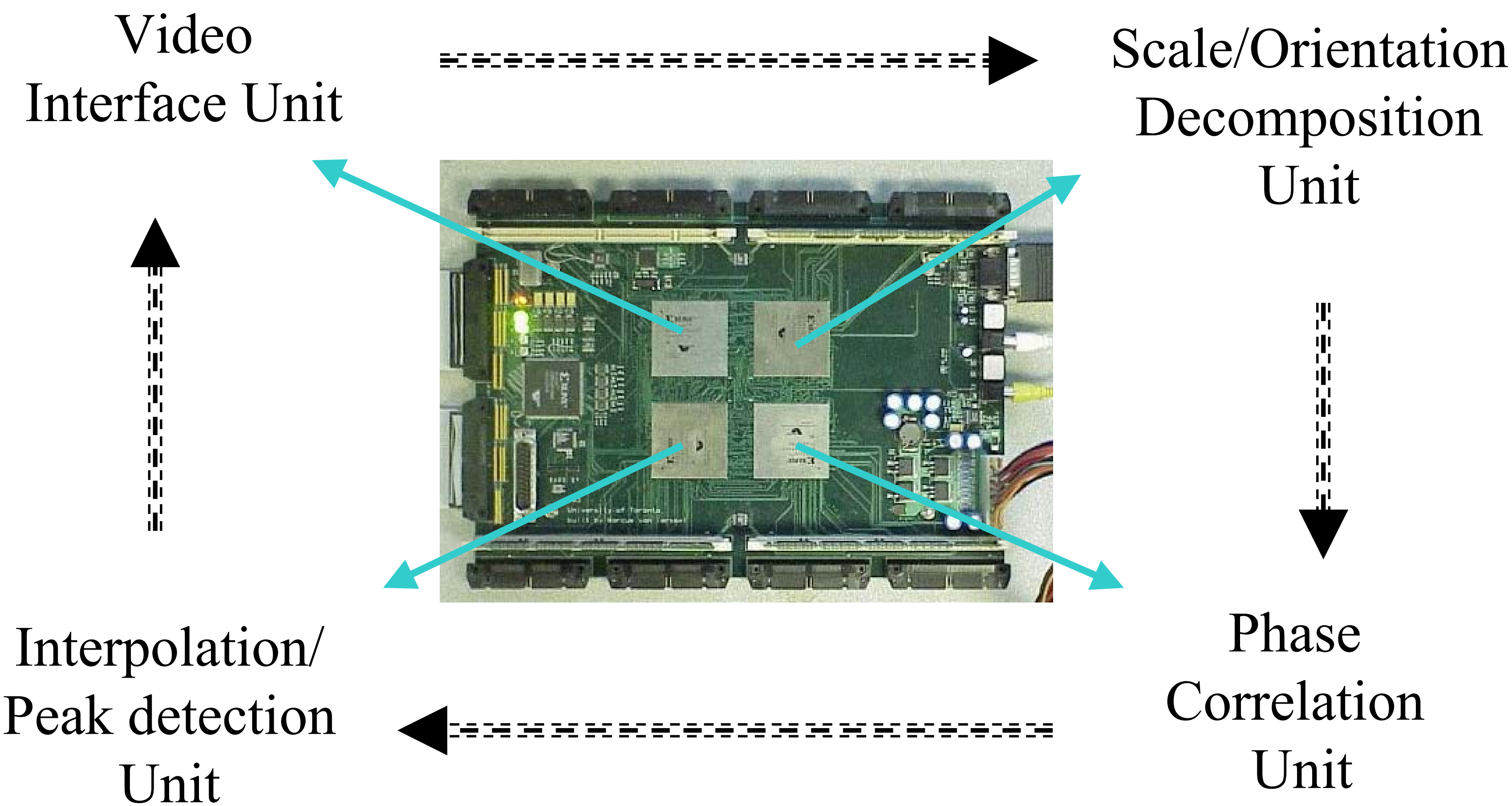
Transmogrifier-3A System



- Four interconnected Xilinx Virtex 2000E FPGAs
- Four external SRAM memory banks
- NTSC/VGA Video ports
- Four general I/O ports

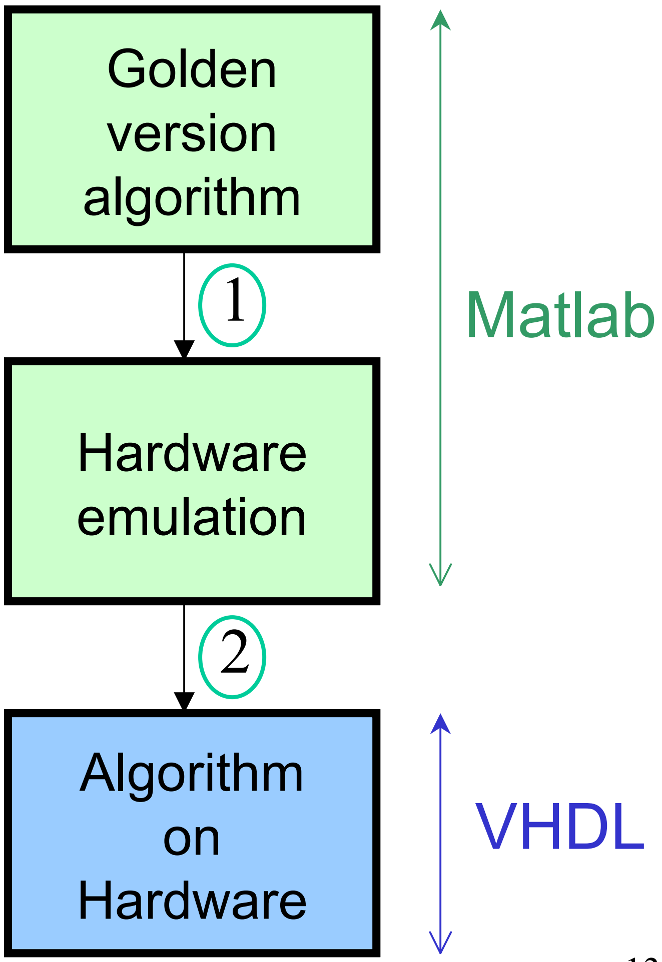
TM-3A system designed in UofT FPGA group

Design Overview



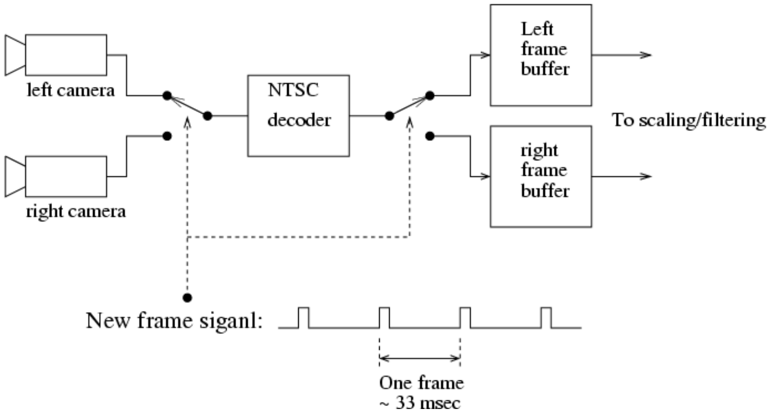
Design Methodology

- Two design steps:
  1. Emulate hardware functional behaviour in software
  2. Build the hardware based on the emulation version



Video Interface Unit

- Input from two cameras in alternating frames
- Output the original image to the display
- Output the depth map results to the display



Scale/Orientation Decompositon Unit

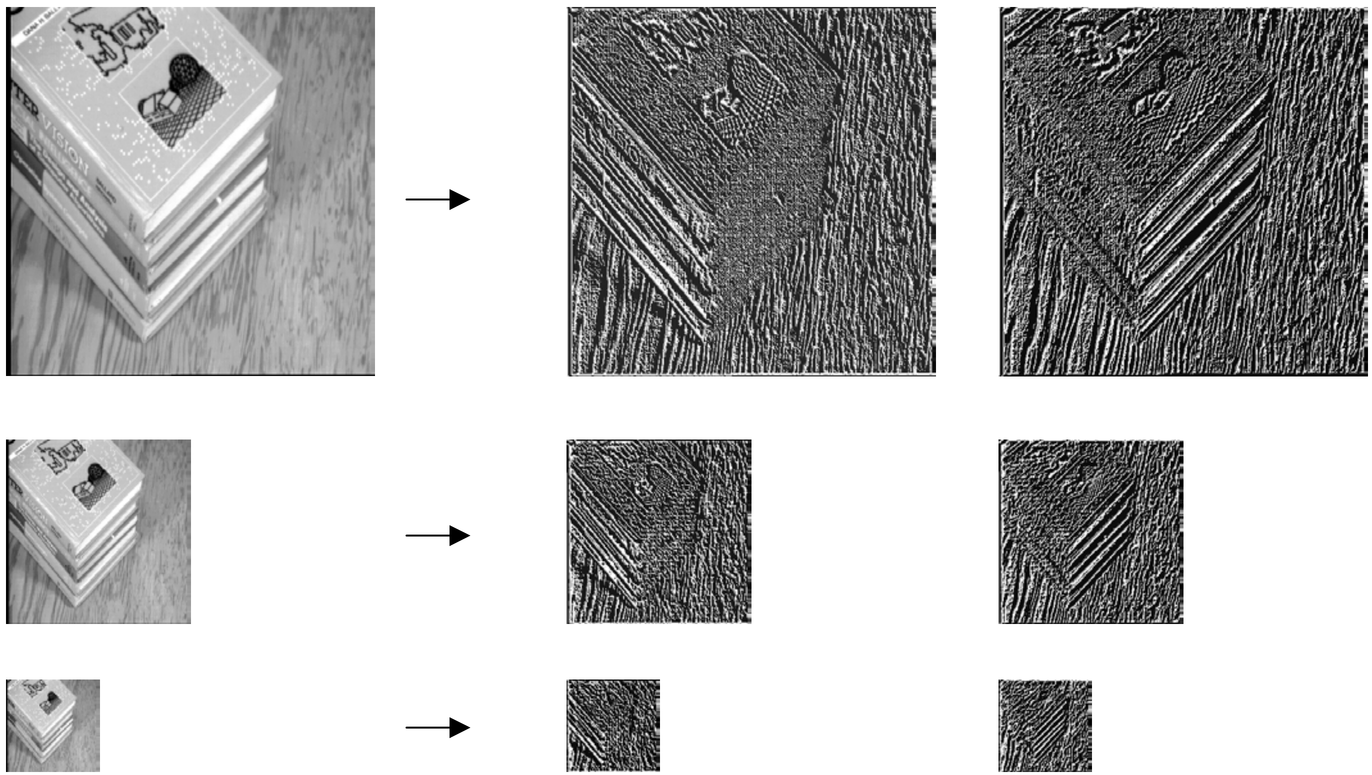
Response Phase (+45 degree)

Response Phase (-45 degree)

Scale 1

Scale 2

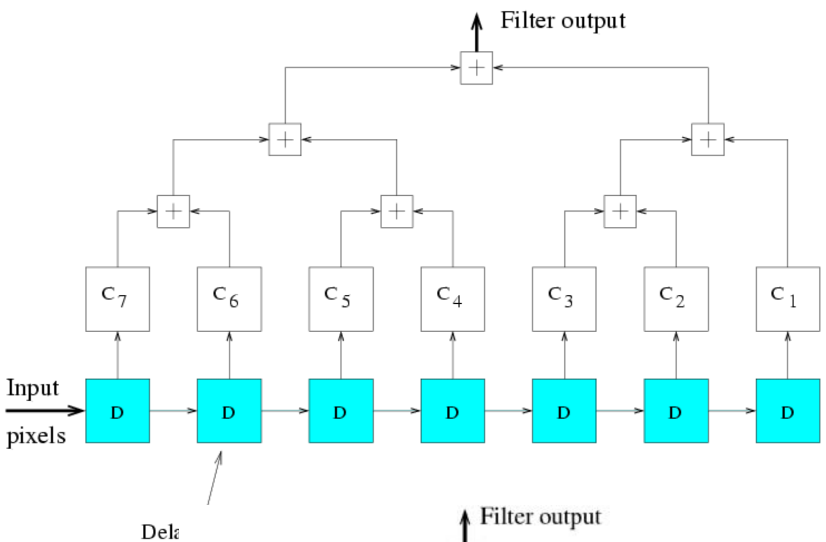
Scale 4

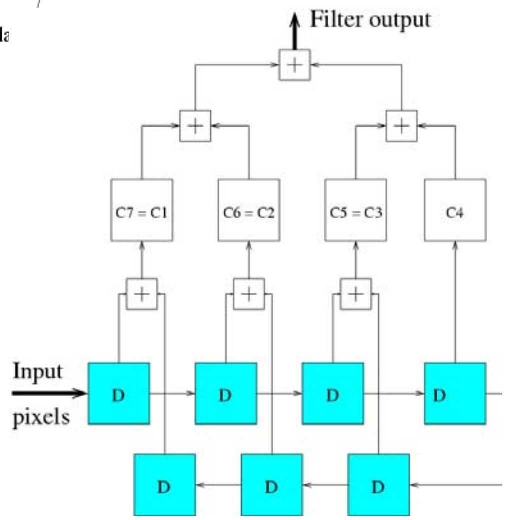


Filtering

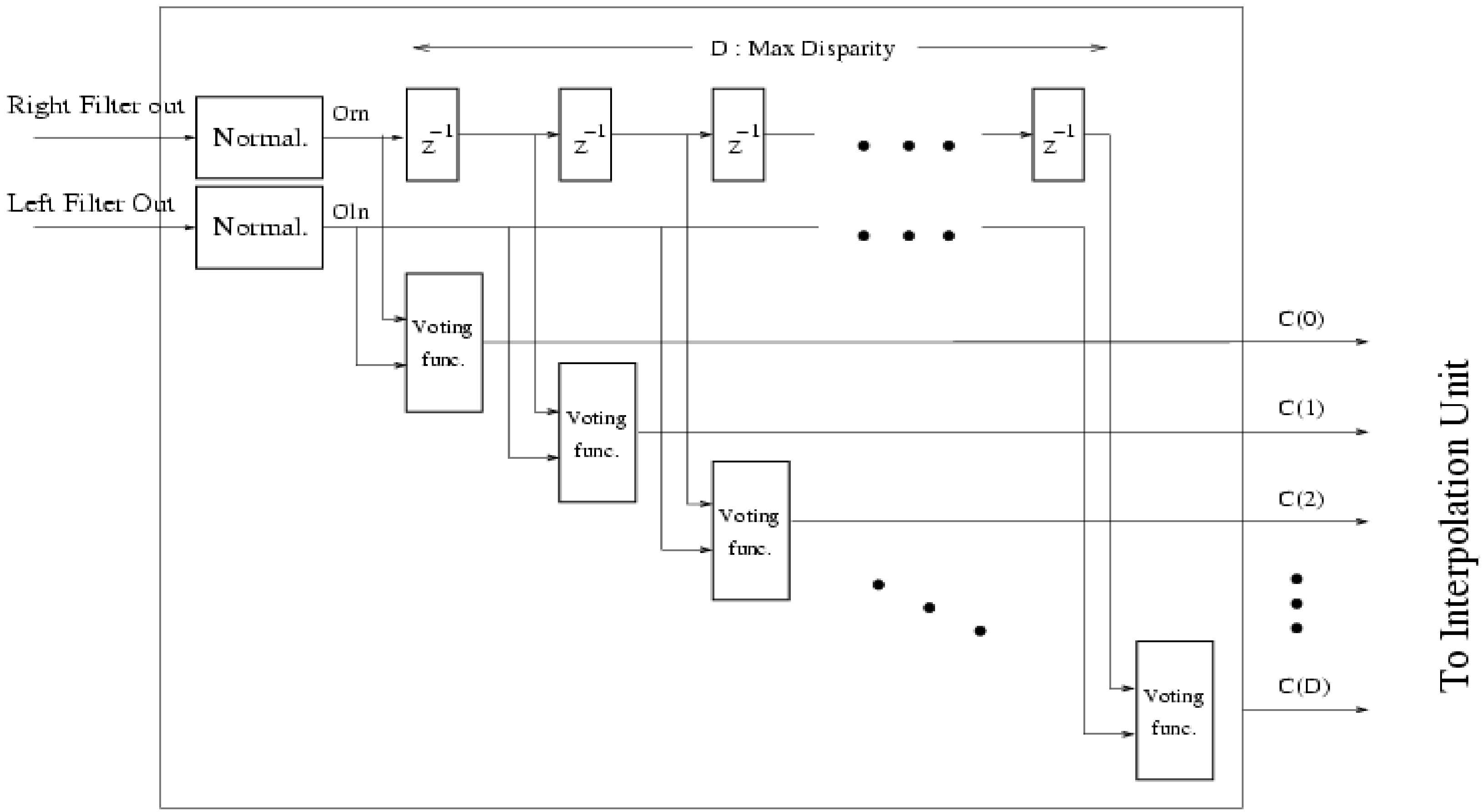
G2/H2 Filters are:

- X_Y separable
  - $O(n^2)$ operations become $O(2n)$
- Symmetrical
  - Reduces # of constant multipliers to half





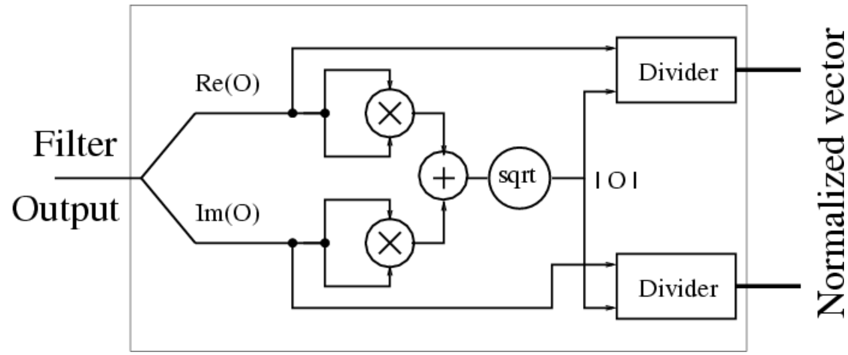
Phase-Correlation Unit



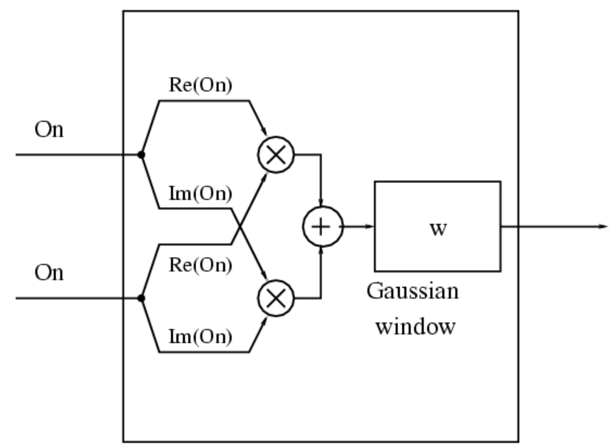
- Left and right images merged

Phase-Correlation Unit

- Normalization block shared for all voting blocks



- Voting block only 2 Multipliers, one adder and one Gaussian window

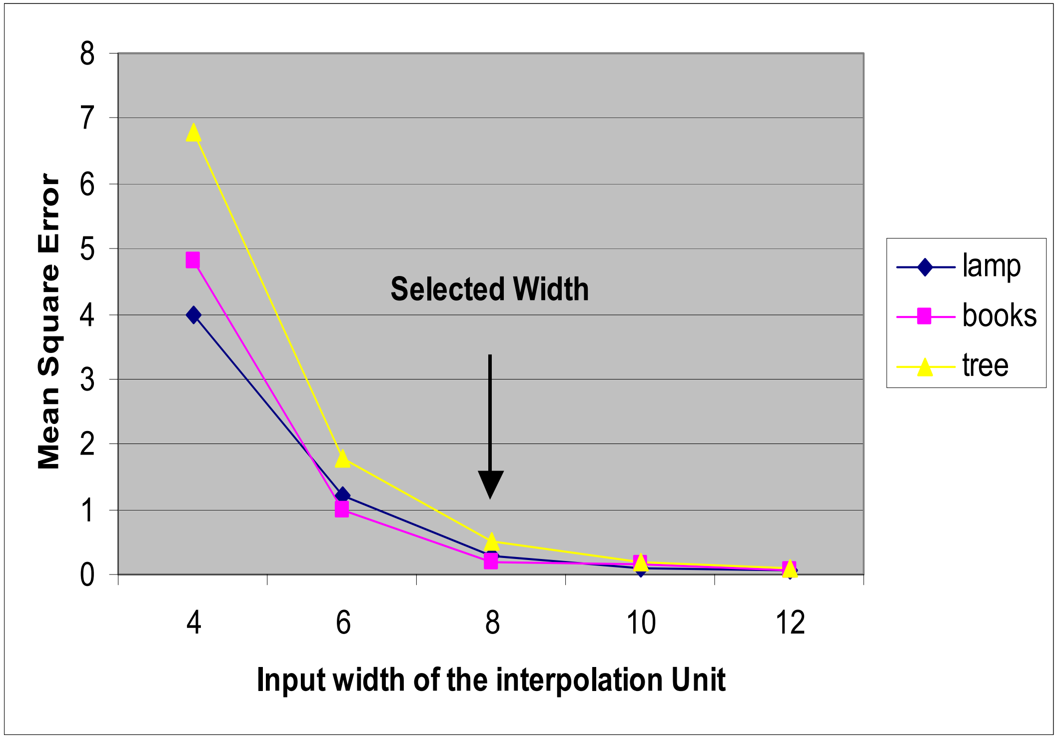


Interpolation/Peak detection Unit

- Combine the voting results over all scales
- Detect the index for the peak value in the overall voting result
- Sub-pixel accuracy
  - fitting the the maximum value and its neighbours to a quadratic curve
  - Accuracy improved from 5 bits to 8 bits

Floating-point to fixed-point conversion

- Fixed-point operations required for efficient implementation
- Analysis is done for every stage
- Efficient enough for our system



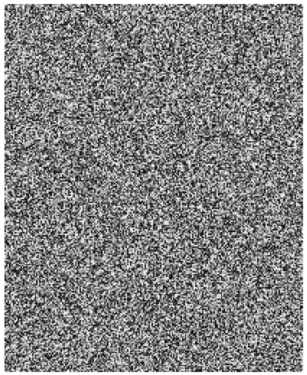
Results

system	m x n (pix.)	D (pix.)	T (msec)	PDS (million)	Algorithm	platform
INRIA	256 x 256	32	280	7.5	Intensity correlation	23 Xilinx XC3090
PARTS	240 x 320	24	23.8	77	Census	16 Xilinx 4025
CMU	200 x 200	30	33	36	Sum of abs. difference	custom hardware
This Work	256 x 360	20	33	55	LWPC	4 Xilinx V2000E

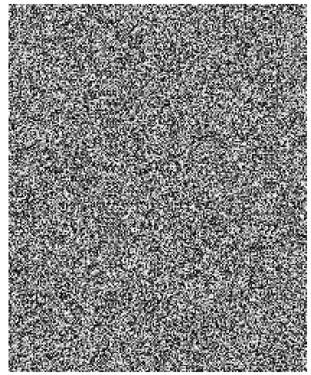
$$PDS = m.n.D / T$$

m x n : Image Size (pixels)
D : Maximum disparity (pixels)
T : Total time for each frame

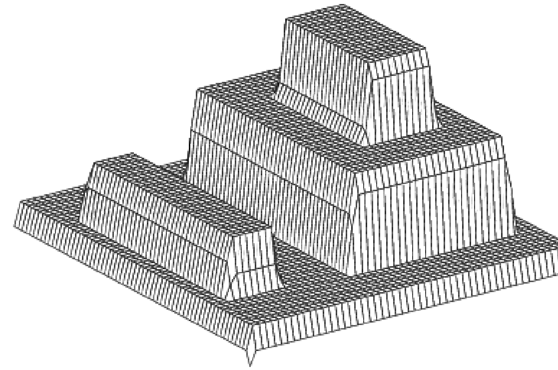
Results: Random Stereograms



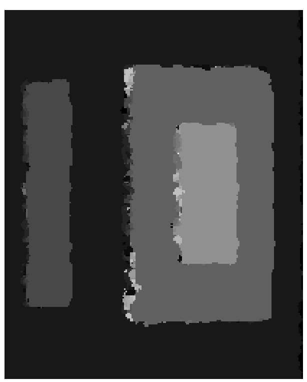
left



right

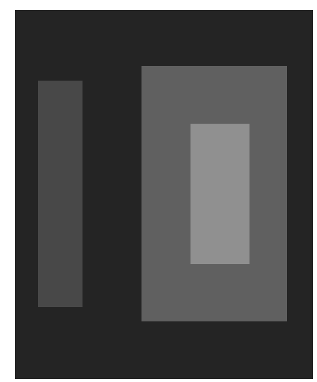


Ground Truth (3D)



Original
Software

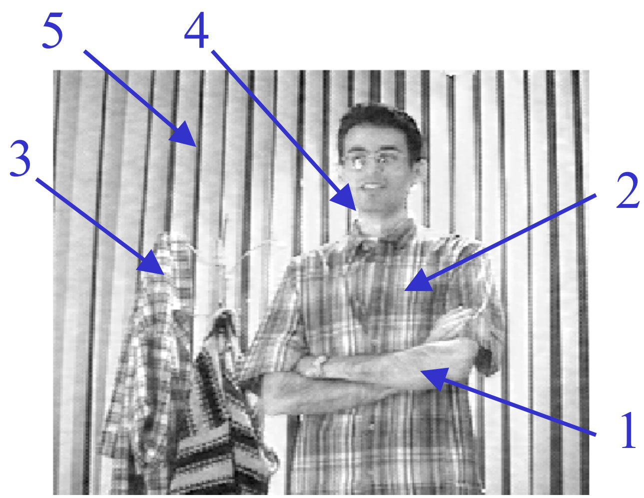
Hardware



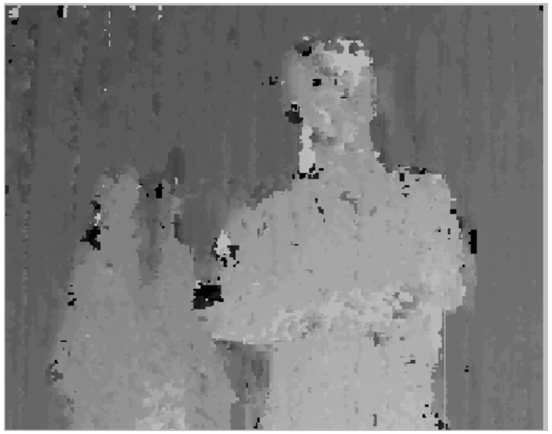
Ground Truth
Depth amp

Results: Natural Images

Left input

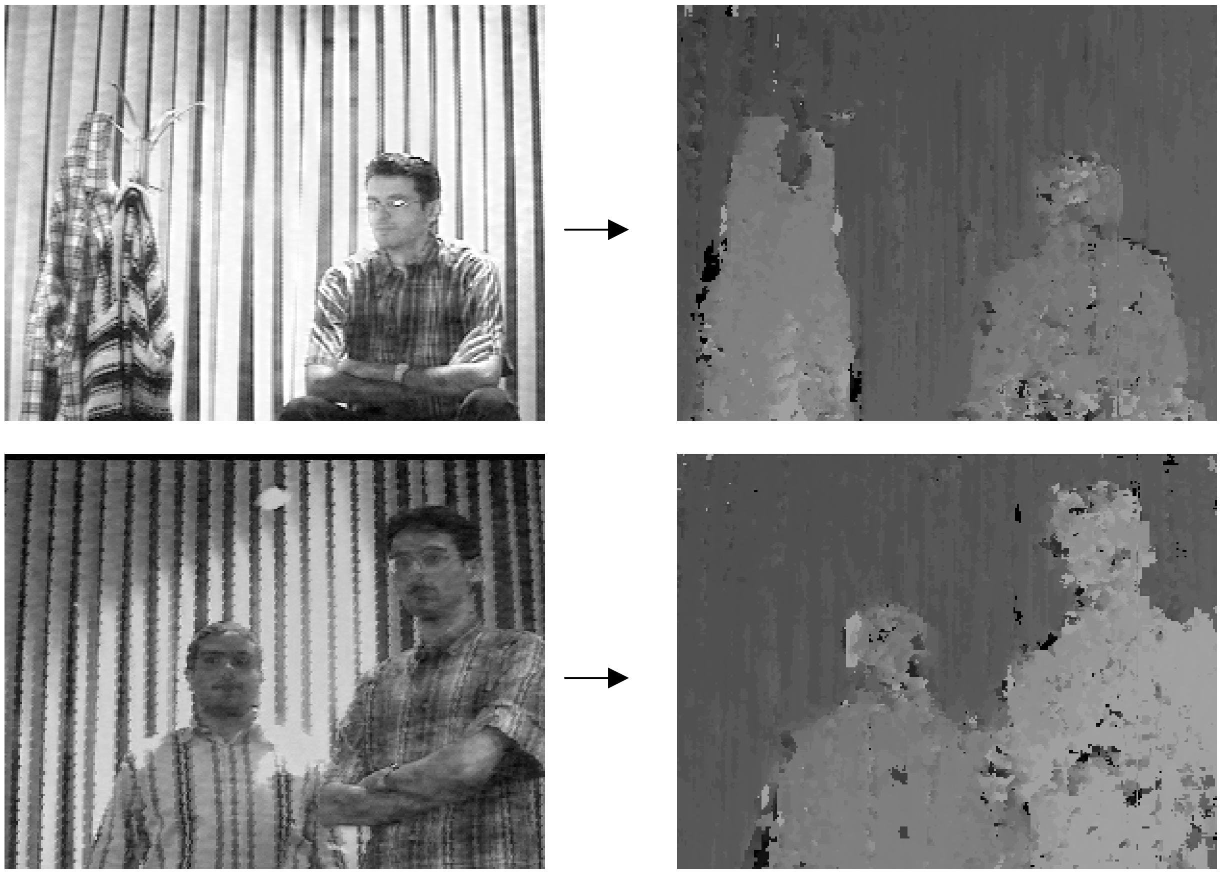


Depth map from hardware



Point #	Ground Truth distance (cm)	hardware results (cm)	% Error
1	300	309	3%
2	315	320	1.6%
3	320	276	13.7%
4	365	355	2.7%
5	410	402	1.9%

More Results



input

depth map from hardware

Conclusion

- Video rate performance (30 frames/sec)
- High accuracy phase-based stereo matching algorithm
- Reprogrammability allows design expansions with minimum cost

Future Work

- extensions to this system:
    - Post-processing blocks to validate the results
    - Using depth information from previous frame
    - Pre-processing blocks to rectify the images
    - Increase the search window size
    - Processing larger images
- Other vision algorithms
- Design automation tools