Frame-rate stereopsis using non-parametric transforms and programmable logic

Peter I. Corke 1 Paul A. Dunn 1 Jasmine E. Banks 23

CSIRO Division of Manufacturing Technology 1
CRC for Mining Technology and Equipment 2
Space Centre for Satellite Navigation 3
Pinjarra Hills, AUSTRALIA 4069.
http://www.cat.csiro.au/cmst/automation

Abstract

A frame-rate stereo vision system, based on non-parametric matching metrics, is described. Traditional metrics, such as normalized cross-correlation, are expensive in terms of logic. Non-parametric measures require only simple, parallelizable, functions such as comparators, counters and exclusive-or, and are thus very well suited to implementation in reprogrammable logic.

1 Introduction

Reliable and real-time 3D perception is critical to the operation of robots in complex real-world environments. This paper discusses recent developments in matching algorithms and in computing technology which together can meet this need.

In our work we have concentrated on area-based stereo matching since it provides dense range maps and works well in the natural outdoor scenes we are interested in. Stereo vision requires no special scene illumination or moving parts, and the sensors (CCD cameras) are small, low-cost, rugged and can acquire data very rapidly. The primary disadvantage has always been that considerable computation is required to establish correspondence or matching of points between the two images.

A number of software only solutions are now available commercially, for instance the Triclops[1] and SVM[2] modules. Using modern Pentium processors, and exploiting the MMX SIMD architecture extension very high matching rates can be achieved. However in order to achieve multiple frames per second even these systems have to work with very low resolution images.

The capability and performance of some reported systems are compared in Table 2.

Section 2 discusses area-based matching metrics: classical as well as non-parametric transforms. Section 3 discusses a reprogrammable logic architecture, and Section 4 describes how that it is applied to frame-rate stereo matching.

2 Area-based matching

At the heart of all area-based matching systems is a matching metric, and a number of classical metrics are listed in Table 1. These compute similarity between square regions of pixels $I_1$ and $I_2$.

Non-parametric techniques[3] are based on the relative ordering of pixel intensities within a window, rather than the intensity values themselves. Consequently, these techniques are robust with respect to radiometric distortion, since differences in gain and bias between two images will not affect the ordering of pixels within a window. In addition they are tolerant to a small number of outliers within a window, and are therefore robust with respect to small amounts of random noise[4]. Two non-parametric transforms which are suited to fast implementation are[3]:

Rank Transform This is defined as the number of pixels in the window whose value is less than the centre pixel. The images will therefore be transformed into an array of integers, whose value ranges from 0 to $N - 1$, where $N$ is the number of pixels in the window.

Census Transform This transform maps the window surrounding the centre pixel to a bit string. If a particular pixel's value is less than the centre
Table 1: Classical area-based matching measures.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sum of Absolute Differences</td>
<td>$\sum_{(u,v)\in W}</td>
</tr>
<tr>
<td>Zero mean Sum of Absolute Differences</td>
<td>$\sum_{(u,v)\in W} \left</td>
</tr>
<tr>
<td>Sum of Squared Differences</td>
<td>$\sum_{(u,v)\in W} (I_1(u,v) - I_2(x+u,y+v))^2$</td>
</tr>
<tr>
<td>Zero mean Sum of Squared Differences</td>
<td>$\sum_{(u,v)\in W} \left( (I_1(u,v) - \bar{I}_1) - (I_2(x+u,y+v) - \bar{I}_2) \right)^2$</td>
</tr>
<tr>
<td>Normalised Cross Correlation</td>
<td>$\sqrt{\sum_{(u,v)\in W} I_1^2(u,v) \cdot \sum_{(u,v)\in W} I_2^2(x+u,y+v)}$</td>
</tr>
<tr>
<td>Zero mean Normalised Cross Correlation</td>
<td>$\sqrt{\sum_{(u,v)\in W} (I_1(u,v) - \bar{I}<em>1)^2 \cdot \sum</em>{(u,v)\in W} (I_2(x+u,y+v) - \bar{I}_2)^2}$</td>
</tr>
</tbody>
</table>

Figure 1: Overall matching process using non-parametric transform.

The steps involved in matching using the rank and census transforms are shown in Figure 1. The stereo images are first transformed then matched using a traditional measure. We choose the SAD metric where the difference is arithmetic for the rank transform, and the Hamming distance

$$\sum_{(u,v)\in W} \text{Hamming}(I_1'(u,v), I_2'(x+u,y+v))$$

for the census transform.

2.1 Results

The stereo pair, Figure 2[9], suffers significant radiometric distortion, that is, one image is brighter than the other. The disparity results obtained using the matching metrics above are shown in Figure 2. In each case, the disparity map is with respect to the right image and lighter regions correspond to larger disparities. A matching window size of 11 x 11 was used for each metric. The census and rank transforms were performed using windows of size 5 x 5.

It can be seen that the SAD and the SSD are not robust with respect to radiometric distortion. Use of the ZSAD, ZSSD, NCC and ZNCC resulted in improved performance but at the cost of increased computational complexity. The NCC and ZNCC are particularly expensive due to the presence of floating point multiplication, division and square root operations. More quantitative results are given in Banks et al[10].

3 The Configurable Logic Processor

A configurable logic device is a digital integrated circuit which contains combinatorial logic (logic gates) and state devices (flip-flops and RAM) which can be interconnected to form complex circuits. The interconnections between these components is by means of switches controlled by the contents of static memory.
Figure 2: Disparity of IROCKS1 stereo pair, produced using various metrics.
on the chip. By the simple means of loading the configuration memory with an appropriate data pattern, arbitrary logical circuits are created. These devices are known as Field Programmable Gate Arrays (FP-GAs), the largest of which exceed 100,000 gates and 10,000 flip-flops.

An FPGA can be used as a computing device by constructing digital circuits which directly transform the data. The use of digital circuits specifically designed for the task at hand has of course been widespread, however until the advent of the FPGA this was an exercise whose cost had to be amortised over a large number of identical units to be worth considering. Application-specific logic has the principal advantages that

- the logic is specifically tailored to the task so that registers, adders, data paths etc. need be no larger that the data width required
- only functional units specifically required need be included, since the hardware is not general
- there is no sequential fetching and decoding of instructions, so hardware for this is not needed
- arbitrary degrees of parallelism are possible within the constraints of available resources and the disadvantages are that the:
  - interconnect tracks and switching, and configuration memory, occupy a substantial chip area and therefore the logic density is much lower than comparable logic devices
  - switched logic paths result in slower propagation of signals

Processing clock rates up to 50MHz can be employed with current FP-GAs, but this is rather slower than the clock rates of 300MHz and more that we now see in conventional CPUs. Effective use of FP-GAs for computation is warranted where the algorithm meets the following conditions:

- the parallelism that can be employed is at least 10
- the arithmetic requirements are simple

The parallelism that can be achieved is often surprising as there are usually many housekeeping tasks that can be done in parallel with data operations. Parallelism can also be increased substantially by pipeline operations. In image processing the data is usually 8 bits and fixed point operations can be used, so arithmetic complexity is moderate.

The CLP[11], see Figure 3, was developed in 1993 and is a VMEbus circuit board containing several FP-GAs which can be configured via the VMEbus. The board includes digital input and output ports and associated timing compatible with the Datacube 10 MHz MAXBUS digital video format.

---

Table 2: Summary of reported stereo vision system performance. L/R indicates left right consistency checking, figure of merit FOM = T/ndm and is an indication of the time to perform a single similarity measure (small is good). n x m is the image size and d is the number of disparities computed.

<table>
<thead>
<tr>
<th>Group</th>
<th>n x m</th>
<th>d</th>
<th>T</th>
<th>FOM</th>
<th>L/R</th>
<th>comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMU iWarp[5]</td>
<td>240 x 256</td>
<td>16</td>
<td>64 ms</td>
<td>65</td>
<td>yes</td>
<td>64 processor iWarp computer</td>
</tr>
<tr>
<td>CMU STORM</td>
<td>256 x 256</td>
<td>32</td>
<td>2.82</td>
<td>1370</td>
<td></td>
<td>Sparc 10</td>
</tr>
<tr>
<td>CMU hardware[6]</td>
<td>256 x 256</td>
<td>32</td>
<td>59</td>
<td>28000</td>
<td>yes</td>
<td>Sparc 2, forward-back match</td>
</tr>
<tr>
<td>CMU DSP[6]</td>
<td>256 x 256</td>
<td>32</td>
<td>9.6</td>
<td>4600</td>
<td>yes</td>
<td>4 M96002 DSP</td>
</tr>
<tr>
<td>CMU software[6]</td>
<td>256 x 256</td>
<td>32</td>
<td>0.28</td>
<td>134</td>
<td>yes</td>
<td>PeRLc-1 board</td>
</tr>
<tr>
<td>JPL[7]</td>
<td>64 x 60</td>
<td>32</td>
<td>0.6</td>
<td>1630</td>
<td>yes</td>
<td>Datacube + 68040</td>
</tr>
<tr>
<td>CMST CLP</td>
<td>256 x 256</td>
<td>32</td>
<td>34 ms</td>
<td>16</td>
<td>no</td>
<td>custom FPGA system†</td>
</tr>
<tr>
<td>PARTS engine[8]</td>
<td>320 x 240</td>
<td>24</td>
<td>23.8 ms</td>
<td>13</td>
<td>no</td>
<td>custom FPGA system</td>
</tr>
<tr>
<td>SRI SVM</td>
<td>100 x 120</td>
<td>32</td>
<td>45 Hz</td>
<td>36</td>
<td>yes</td>
<td>233 MHzPentium+MMX</td>
</tr>
<tr>
<td>Triclops</td>
<td>160 x 120</td>
<td>16</td>
<td>10 Hz</td>
<td>325</td>
<td>yes</td>
<td>Trinocular 266 MHzPentium+MMX</td>
</tr>
</tbody>
</table>

---

1Xilinx 4013 each of 13,000 'gates'.
The board contains four 512 x 512 8-bit frame stores which can be switched between various FPGAs. These use 20 ns static RAMs and are connected via a shuffle network to four of the FPGAs. All address, data and control signals are separate so as to allow independent access to each frame store. There are two 8,000 gate FPGAs for managing image input and output and four 13,000 gate FPGAs for data processing.

4 Configurable Logic Implementation

Our implementation has three major stages:

1. Compute census values over all 5 x 5 windows in each image.

2. Compare two census values using the exclusive OR operation and count the number of zero bits resulting. The sum of the number of matching bits over a larger, 11 x 11, window provides the similarity measure.

3. Find the maximum of the array of computed similarity measures, the index being the disparity.

Our stereo-matcher is implemented for an image size of 256 x 256 and two CLP boards are required.

4.1 Programming the FPGAs

Two compilers\(^2\) have been developed for FPGA programming. A circuit generation language, \(\text{http://www.mlb.dmt.csiro.au/IVT/clp/clp/clpuse.html}\)

\(^2\)http://www.mlb.dmt.csiro.au/IVT/clp/clp/clpuse.html

Figure 4: Data flow through the FPGA system.

Figure 5: Timing diagram for the CLP implementation.
CCGL[11], was developed first and is based on C language syntax. Its output is a logic netlist which is passed to the proprietry Xilinx tools for logic placement and routing. CCGL was developed before VHDL was available and continues to be used. It has proved to be an enormously flexible and concise way of expressing designs while still allowing optional control of placement and timing. Unfortunately such hardware description languages still require the user to be accomplished in circuit design.

4.2 Next generation CLP

The CLP board is no longer state-of-the-art and development of a more ambitious configurable logic processor is nearly complete. The new system, A Hybrid Modular Processor, provides three building blocks:

- an FPGA module for primary parallel logic
- a DSP module for intensive arithmetic or complicated sequential code
- an I/O module to interface to cameras and other external hardware devices

A selection of these modules can be interconnected into a processing network of a size and topology appropriate to the application.

Each module has two input and two output 20-bit data ports with a data rate of 60Mword/s. Data port interconnections are simple coaxial cables carrying serial data at 1.2GHz. As the processing network grows the data communications grows with it and hence there are no bottlenecks. Module control and communications is accomplished via an IEEE1394 (Firewire) backbone to a host processor (PC or workstation). Each module also carries a DEC/Intel StrongARM processor to manage backbone communications and provide any extra local processing.

The FPGA module contains 9 completely independent frame stores, 2 large lookup tables, a Xilinx 28,000 gate FPGA for input data buffering and control and uses a Xilinx 0.9 ns 150,000 gate array with 40 MHz, 60 MHz and 80 MHz clocks for processing. The DSP module contains a TI TMS320C6701. The I/O module couples the StrongARM processor with an FPGA for handling a diverse range of cameras, analog and signal conversion being managed via mezzanine board.

5 Conclusions

This paper has briefly discussed non-parametric transforms for stereo matching. These transforms exhibit matching quality at least as good as more traditional matching measures and have the significant advantage of being ammenable to hardware implementation. A frame-rate hardware implementation based on reconfigurable logic was briefly described. Increasing density and speed of FPGA technology would allow the complete system to now be implemented in a single chip.

References