A Generic Implementation Framework for Stereo Matching Algorithms

Jasmine Banks[1,2], Reid Porter[1], Mohammed Bennamoun[1] and Peter Corke[2,3]

[1]Space Centre for Satellite Navigation,
Queensland University of Technology.
[2]Cooperative Research Centre for Mining Technology and Equipment.
[3]CSIRO Manufacturing Science and Technology.
Email: {j.banks,r.porter,m.bennamoun}@qut.edu.au, pic@cat.csiro.au

Abstract

Traditional area-based matching techniques make use of similarity metrics such as the Sum of Absolute Differences (SAD), Sum of Squared Differences (SSD) and Normalised Cross Correlation (NCC). Non-parametric matching algorithms such as the rank and census rely on the relative ordering of pixel values rather than the pixels themselves as a similarity measure. Both traditional area-based and non-parametric stereo matching techniques have an algorithmic structure which is amenable to fast hardware realisation. This investigation undertakes a performance assessment of these two families of algorithms for robustness to radiometric distortion and random noise. A generic implementation framework is presented for the stereo matching problem and the relative hardware requirements for the various metrics investigated.

Keywords: *stereo vision, image matching, rank transform, census transform, implementation*

1 Introduction

A fundamental problem in stereo vision is that of locating *corresponding* or *matching* points in the two images. *Area-based* matching algorithms are characterised by the fact that they compare actual grey-level pixel values in the two images in order to find the best match. Usually regularly sized pixel neighbourhoods are compared, since the grey-level information contained in a single pixel is insufficient for unambiguous matching. Area-based techniques are well suited to textured surfaces[11], and have the potential to yield a dense depth map[9]. In addition, they have an algorithmic structure amenable to fast hardware implementation[8].

Section 2 outlines the principle of area-based matching and describes a few commonly used matching measures. Section 3 then discusses non-parametric transforms, in particular, the rank and the census transform. Experimental results obtained using both traditional area-based metrics and non-parametric transforms are shown in Section 4. Issues associated with the relative hardware requirements of these techniques are then discussed in Section 5.

2 Area-Based Matching

In area-based matching, a point to be matched essentially becomes the centre of a small window of pixels, and this window is compared with similarly sized regions in the other image. *Matching metrics* are used to provide a numerical measure of the similarity between a window of pixels in one image and a window in another image, and hence are used to determine the optimum match.

Epipolar geometry[4] is used to improve the efficiency of the matching process by constraining the search to one dimension. Stereo images may be rectified such that the epipolar lines correspond to the horizontal scan lines[2]. A

Sum of Absolute Differences	SAD	$\sum_{(u,v)\in W} \lvert I_1(u,v) - I_2(x+u,y+v)\rvert$
Zero mean Sum of Absolute Differences	ZSAD	$\sum_{(u,v)\in W} \left\lvert (I_1(u,v) - \overline{I_1}) - (I_2(x+u,y+v) - \overline{I_2})\right\rvert$
Sum of Squared Differences	SSD	$\sum_{(u,v)\in W} (I_1(u,v) - I_2(x+u,y+v))^2$
Zero mean Sum of Squared Differences	ZSSD	$\sum_{(u,v)\in W} ((I_1(u,v) - \overline{I_1}) - (I_2(x+u,y+v) - \overline{I_2}))^2$
Normalised Cross Correlation	NCC	$\frac{\sum_{(u,v)\in W} I_1(u,v)\cdot I_2(x+u,y+v)}{\sqrt{\sum_{(u,v)\in W} I_1^2(u,v)\cdot \sum_{(u,v)\in W} I_2^2(x+u,y+v)}}$
Zero mean Normalised Cross Correlation	ZNCC	$\frac{\sum_{(u,v)\in W} (I_1(u,v) - \overline{I_1})\cdot (I_2(x+u,y+v) - \overline{I_2})}{\sqrt{\sum_{(u,v)\in W} (I_1(u,v) - \overline{I_1})^2 \cdot \sum_{(u,v)\in W} (I_2(x+u,y+v) - \overline{I_2})^2}}$

Table 1: Area based matching measures[1]. In all cases, I_1 denotes the template window, I_2 is the candidate window, and $\sum_{(u,v)\in W}$ indicates summation over the window.

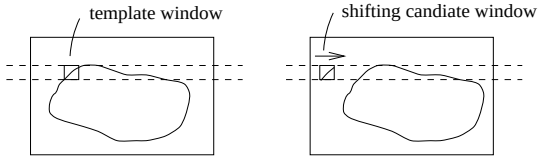

Figure 1: Epipolar constrained area based matching.

simple approach used in area based matching is to compute the value of the matching metric using a fixed window in the first image and a shifting window in the second image, as illustrated in Figure 2. The shifting window is moved in integer increments along the epipolar line, where the amount of shift is the test disparity[9]. The disparity having the optimum value for the matching metric is then chosen. A number of metrics which use a square window of pixels as the basis for comparison are listed in Table 1.

The SAD and the SSD are the simplest, and computationally the least expensive of all the matching measures. Two areas which consist of exactly the same pixel values would yield a score of zero. However, these measures will no longer yield the correct results in the case of radiometric distortion, ie, where the pixel values in one image differ from those in the other image by a constant offset and/or gain factor[10]. The ZSAD and the ZSSD have been devised to deal with this problem, by subtracting the mean of the match area from each intensity value. However, the improved performance of the ZSAD and ZSSD over the SAD and SSD is offset by substantially increased computational complexity. The NCC measure deals with a possible gain factor by dividing by the variances of each window, while the ZNCC measure additionally deals with the offset problem by first subtracting the mean from each pixel value. These metrics will have a value ranging from -1 to 1, where 1 represents the best match.

3 Non-Parametric Techniques

Non-parametric techniques 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[3]. In addition, these transforms are tolerant to a small number of outliers within a window, and are therefore robust with respect to small amounts of random noise[5].

Two non-parametric transforms which are suited to fast implementation are the rank transform and the census transform[13].

3.1 Rank Transform

The rank transform is defined as the number of pixels within a 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. A pair of rank transformed images are then matched using one of the matching metrics of Section 2. For hardware implementation, it is advantageous to use a matching metric based on integer arithmetic, such as the SAD or the SSD.

3.2 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 pixel then the corresponding position in the bit string will be set to 1, otherwise it is set to zero. Two census transformed images are compared using a similarity metric based on the Hamming distance, ie, the number of bits that differ in the two bit strings. The Hamming distance is summed over the window, ie,

$$\sum_{(u,v)\in W} Hamming(I_1'(u,v), I_2'(x+u, y+v)) \quad (1)$$

where I_1' and I_2' represent the census transforms of I_1 and I_2. Two hardware implementations of this scheme are discussed in [7, 12].

4 Experimental Results

The area-based matching metrics of Table 1 and the rank and census transforms were implemented in software, in order to test their performance with a number of test images. Each algorithm accepts a rectified stereo pair as input and produces a disparity map as output. In each case, the left-right consistency criterion[9] was applied, in order to remove invalid matches. In addition, isolated matches which remain after left-right checking were removed, based on the assumption that these matches are likely to be incorrect[8, 9].

The disparity results obtained for the J1 stereo pair of Figure 2 are shown in Figure 3, where lighter regions in the disparity maps correspond to larger disparities. A matching window of size 11×11 was used in each case. The J1 test pair was among those used in the JISCT stereo evaluation[6]. It is noticeably affected by radiometric distortion, the right image being approximately 13% brighter than the left.

5 Real Time Implementation

The real time implementation requirements of the similarity metrics of Table 2 were investigated with the hardware description language VHDL. For the purposes of comparison the algorithms were implemented with full precision fixed point arithmetic. The main components of the devised stereo matching system are illustrated in Figure 4. The most important design constraint was the memory bandwidth. This is a common problem in window based image processing due to the difficulty in accessing a local neighbourhood in linear memory. A pipelined architecture was adopted that utilized local memory and shift registers where possible. The inherent parallelism of the stereo matching algorithms was exploited by calculating the different disparities in parallel.

5.1 Transform

The transform components depicted in Figure 4 are only required in the case of the rank and census transforms. The image transform was required to supply the transform neighbourhood at the pixel rate. The transforms were implemented for a 5×5 neighbourhood using 24 comparators operating in parallel.

5.2 Point Operator

In real time robotic applications disparities are calculated for every point in an image pair. With this in mind the idea of a "combined image" is useful. A combined image is produced by application of the algorithm point operation to the left and right images. The maximum disparity, D, dictates the number of combined images that are produced. The images are offset horizontally for each disparity value before being combined. A series of delay elements were used on one of the image pixel streams to generate the required horizontal offsets associated with each disparity, as illustrated in Figure 4. The point operations for the different similarity metrics are summarized in Table 2.

5.3 Window Summation

Once the combined images have been calculated the matching window must be summed. The window overlap between adjacent pixels means that there are redundant additions. These can be avoided by using running totals as suggested in [8]. This can be considered a

Figure 2: Stereo pair J1.

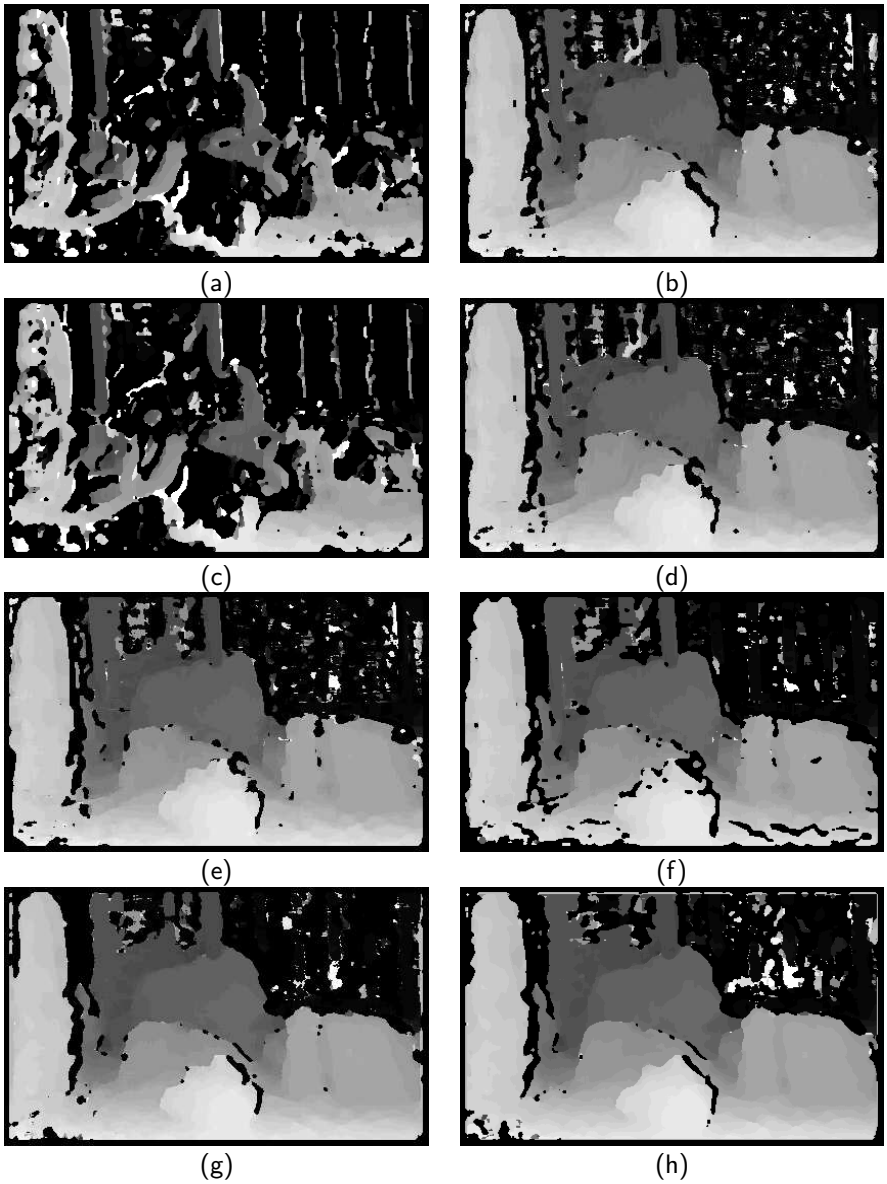

Figure 3: Disparity of J1 stereo pair, produced using (a) SAD, (b) ZSAD, (c) SSD, (d) ZSSD, (e) NCC and (f) ZNCC metrics, as well as (g) Rank transform followed by the SAD and (h) Census transform followed by the Hamming metric.

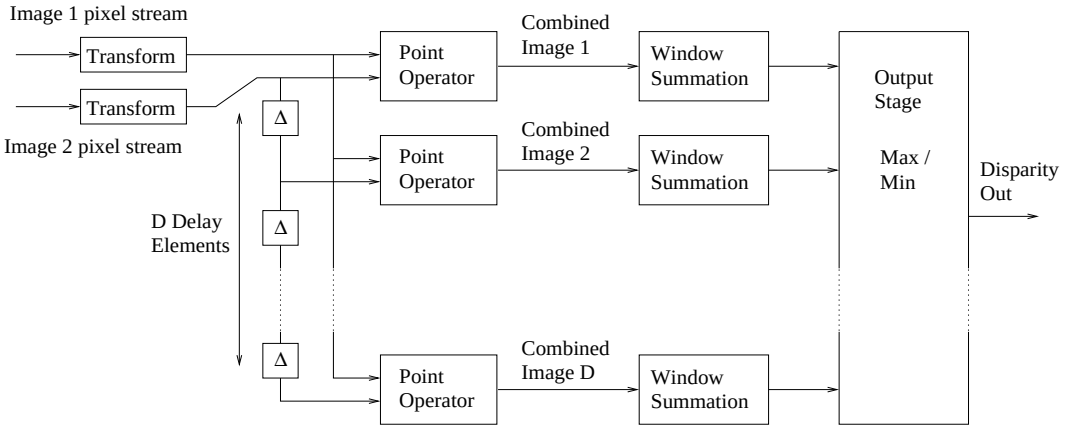

Figure 4: Components of stereo matching system.

Similarity Metric	Point Operation
SAD/ZSAD/Rank	Subtraction
SSD/ZSSD	Subtraction/ Squarer
NCC	Multiplication
Census	XOR/bit count

Table 2: Similarity metric point operations.

two stage process:

1. Calculate row sums : Sum the first window row and store the result. To calculate the next row sum, the next pixel is added to the total and the last pixel is subtracted. This was implemented using a window length shift register and an adder/subtracter. The row sums could be calculated at the pixel rate with ease.

2. Calculate column sums : This is similar to the first step but instead of accumulating pixel values we are accumulating row sums from the previous stage. Since the row sums are being calculated in scan line order, a large number of memory elements are required to subtract the last row sum. There are several ways this can be implemented involving choices between centralized (RAM bank) or local memory (shift registers). An alternative is to introduce redundant additions so that the last row is calculated at the same time as the first row, reducing memory requirements. This was the method used by Paul Dunn and Peter Corke in [7] and the method adopted in our models.

5.4 Output Stage

To produce the final disparity map the window sums from the D combined images are compared so that the optimum match can be chosen. This process is similar for all algorithms under consideration with SAD, SSD, CENSUS and RANK requiring the smallest value to be chosen and NCC the largest.

5.5 Further Considerations

The zero mean versions of the similarity metrics were included in the implementation comparison by assuming image preprocessing. Calculating the zero mean similarity metrics introduces a one frame latency into the design so that frame means can be calculated and deducted before being passed to the image pixel streams in Figure 4.

To include the NCC within the implementation framework some further design issues had to be addressed. The NCC denominator could be calculated either with the same circuit used to calculate the numerator or in parallel by replicating hardware. To keep implementations consistent in time the extra area requirements were used as the comparison measure. The NCC division and square root were implemented using look-up tables.

5.6 Implementation Comparison

Area estimates were calculated for the similarity metrics of Table 2 and are summarised in Table 3. The estimates are based on 8 bit pixel images with a window size of 11×11. These estimates represent a technology independent cost function and have no units.

Similarity Metric	Transform	Point Operator/ D	Window Summation/ D	Further Considerations	Total Estimate D = 32
SAD/ZSAD	ZSAD 400	260	3400		240000
SSD/ZSSD	ZSSD 400	780	5520		405000
NCC		610	5520	6130/D+ LUTs	790000
Census	5600	2400	2600		325000
Rank	8000	190	2600		200000

Table 3: Area cost estimates.

Neighbourhood summation was the predominate cost in terms of area and was dictated largely by the data width. For example the NCC and SSD required 121 sixteen bit numbers to be summed while the transform methods only required 121 five bit numbers. In this respect transform based approaches such as the rank and census were the most efficient in terms of implementation area. This was particularly true for large disparity ranges where the cost of the image transformation was small compared to the total cost of the implementation.

6 Discussion

Figure 3 shows that the SAD and the SSD perform poorly in the case of radiometric distortion. Use of the ZSAD, ZSSD, NCC and ZNCC results in improved robustness with radiometric distortion, however, these metrics introduce increased computational complexity. The implementation area overheads for the zero-mean similarity metrics is small compared to the total area, however a one frame latency is introduced. The NCC and ZNCC are particularly computationally intensive due to the need to calculate the normalising denominator which increases hardware requirements by a factor of two.

Both the rank transform followed by matching with the SAD metric, and the census transform followed by matching with the Hamming metric, were found to be invariant to radiometric distortion, as shown by Figure 3. An additional advantage of both these algorithms is their amenability to fast hardware implementation. Their reduced representation requirements in the window summation means they are prime candidates for use in a real-time stereo matching system.

7 References

[1] P. Aschwanden and W. Guggenbühl, "Experimental Results from a Comparative Study on Correlation-Type Registration Algorithms", Robust Computer Vision, 268–289, Wickmann, 1993.

[2] N. Ayache, Artificial Vision for Mobile Robots, MIT Press, 1991.

[3] J. Banks, M. Bennamoun and P. Corke, "Fast and Robust Stereo Matching Algorithms for Mining Automation", IAIF'97, November 1997.

[4] S. Barnard and M. Fischler, "Computational Stereo", Computing Surveys, Volume 14, Number 4, 553–572, December 1982.

[5] D. Bhat and S. Nayar, "Ordinal Measures for Visual Correspondence", Proceedings of Computer Vision and Pattern Recognition, San-Fransisco, 351–357, 1996.

[6] R. Bolles, H. Baker and M. Hannah, "The JISCT Stereo Evaluation", Image Understanding Workshop, DARPA, 263–274, 1993.

[7] P. Dunn and P. Corke, "Real-time Stereopsis using FPGAs", FPGA97, Imperial College London, September 1997.

[8] O. Faugeras et al, "Real-Time Correlation-Based Stereo: Algorithm, Implementations and Applications", Technical Report 2013, INRIA, 1993.

[9] P. Fua, "A Parallel Stereo Algorithm that Produces Dense Depth Maps and Preserves Image Features", Machine Vision and Applications, Volume 6, 35–49, 1993.

[10] M. Hannah, "Computer Matching of Areas in Stereo Images", PhD thesis, Stanford University, 1974.

[11] M. Hannah, "Digital Stereo Image Matching Techniques", International Archives of Photogrammetry and Remote Sensing, 280–293, 1988.

[12] J. Woodfill and B. Herzen, "Real-Time Stereo Vision on the PARTS Reconfigurable Computer", IEEE Workshop in FPGAs for Custom Computing Machines, 242–250, 1993.

[13] R. Zabih and J. Woodfill, "Non-Parametric Local Transforms for Computing Visual Correspondence", 3rd European Conference on Computer Vision, Stockholm, 1994.