Tina Memo No. 1994-001
Internal.

Tutorial: Overview of Stereo Matching Research.

R.A.Lane and N.A.Thacker.

Last updated
1 / 12 / 1998
See Also Tina Memo: 1995-1998

This document forms part of the **Features and Measurement** Series
available from www.tina-vision.net.

2002-005	Tutorial: An Empirical Design Methodology for the Construction of Machine Vision Systems.
1996-001	Algorithmic Modelling for Performance Evaluation.
2006-002	A Statistical Framework for Detection of Connected Features.
2006-007	Quantitative Verification of Projected Views Using a Power Law Model of Feature Detection.
1997-003	Tutorial: Supervised Neural Networks in Machine Vision.
1995-003	Invariance Network Architecture.
1992-001	Combining the Opinions of Several Early Vision Modules using a Multi-Layer Perceptron.
2005-006	Curve Fitting and Image Potentials: A Unification within the Likelihood Framework.
1996-002	Tutorial: The Likelihood Interpretation of the Kalman Filter.
1994-003	Using a Switchable Model Kalman Filter.
2004-012	Tutorial: Computing 2D and 3D Optical Flow.
2005-011	Comparing the Performance of Least-Squares Estimators: when is GTLS Better than LS?
1994-001	Tutorial: Overview of Stereo Matching Research.
1992-002	Online Stereo Camera Calibration.
1995-002	Calibrating a 4 DOF Stereo Head.
2000-009	An Evaluation of the Performance of RANSAC Algorithms for Stereo Camera Calibration.
2001-011	The Evolution of the TINA Stereo Vision Sub-System.
2007-011	A Methodology for Constructing View-Dependent Wireframe Models.

WWW.TINA-VISION.NET

Imaging Science and Biomedical Engineering Division,
Medical School, University of Manchester,
Stopford Building, Oxford Road,
Manchester, M13 9PT.

1 Overview

Determination of three dimensional data from images is of central importance in the field of machine vision. One of the most direct way of achieving this from image data is stereo vision. Stereo vision has a wide range of potential application areas including; three dimensional map building data visualisation and robot pick and place. A variety of constraints may be used to guide the correspondance solution depending upon the properties of the data. This is reflected in the broad range of algorithms that have been developed. For example, if camera calibration is available epi-polar constraints can be used. The absence of transparent objects allows the use of disparity gradient limits. The absence of occlusion can permit strong surface smoothness constraints. If the images are generated under constrained lighting conditions the images may display photo-metric properties allowing direct pixel matching. All of these factors can have a strong influence on the quantity and reliability of data recovered by an algorithm. Consequently, it is not fair to expect any one algorithm to be capable of making the best of an arbitrary set of data.

In 1988 a survey of 15 institutions [10] observed that, in practice, many researchers were using combinations of the above constraints for solving a wide variety of stereo matching problems. The conclusions from this survey were that each of the different approaches had its relative merits and disadvantages dependent on the nature of the matching problem. This sentiment has also been reiterated by Hannah [13] who suggested that a complete stereo solution would need to combine the relevant approaches in a cooperative fashion. A practical example of a combined approach is perhaps best demonstrated by Baker [2] who successfully combined both edge and luminance based stereo techniques. Bearing this in mind, the following discussion should only be seen as an analysis of the area of applicability of each approach and not as an attempt to rule in or out particular methods.

Throughout this section, a basic knowledge of the common approaches used by stereo matching algorithms is assumed. If this is not the case, then the reader is referred to the following work: [20, 23, 3, 4, 13]. The purpose of this chapter is to analyse the significant pieces of work produced in the area of stereo vision. In order to do this, a categorisation is introduced which loosely divides the work into area based stereo and feature based stereo. Area based stereo is used to classify algorithms which utilise image domain similarity metrics in the correspondence process. Area based algorithms can be further divided in the following categories:

- Cross-correlation based
- Least-squares region growing
- Simulated annealing based

Feature based stereo is defined as algorithms which perform stereo matching with high level parameterisations called image features, these algorithms can be classified by the type of features used in the matching process as follows:

- Edge-string based
- Corner based
- Texture region based

In addition to the different categories listed above, other issues exist which can equally well be applied to the above approaches, namely:

- Human interaction
- Hierarchical processing
- Interpolation of a partial solution

2 Epi-Polar Calibration

The epi-polar constraint is one of the most fundamentally useful pieces of information which can be exploited during stereo matching. It can be shown by elementary geometry [?] that 3D feature points are constrained to lie along epi-polar lines in each image. Knowledge of the epi-polars reduces the correspondance problem to a 1D search. This constraint is best utilised by the process of image rectification. The principle is quite simple, any pair of images can be transformed to a "paralell camera geometry" so that corresponding point features in 3D will lie on the same horizontal line in the two images. Unfortunately such a process requires some knowledge of the left to right camera transformation which will generally require a calibration proceedure [?] The freedom for camera model specification, cost function definition and numerical implementation is immense. It is perhaps somewhat unfortunate that the subject must continue to be a drain on research efforts during the development of new systems. Camera calibration is quite a Pandoras box of publications and methods (the most commonly referenced text is probably [?]) but a few guidelines can be provided as to what constitutes good practice.

The camera model must be specified with a minimum number of parameters which describe the important degrees of freedom. There are at least three ways of representing the left-to-right camera coordinate rotation matrix. The quaternion, the screw (Rodriegez method) and polar co-ordinate triple axis rotation, there is probably very little to distinguish between the performance of these methods. If the cameras display radial distortion effects these may well not be visible to the casual observer but will weaken the accuracy of the epi-polar constraint and completely bias any resulting 3D measurement. Image centres and aspect ratios may also need to be free parameters and can generally not be expected to take default (side of the box) values.

The cost function must be defined in the image plane as the errors between back projected positions for points. This is the only domain in which measurement errors can be expected to be uniform so that systematic errors are not introduced during the calibration process. If the data for calibration is to be obtained from any automatic matching process then the cost function must be in the form of a robust statistic. Any automatic calibration proceedure which is based on least-squares will be susceptible to calibration matching failiures and must be viewed with suspiscion.

A practical calibration system must give some indication of resulting calibration accuracy, either in terms of resulting back projected error or as covariance estimates on the estimated parameters. Ideally any subsequent stereo algortihm should be capable of interpreting these measures and adjusting output depth accuracy estimates accordingly.

A recent focus of research has been in the area of "self-calibration" [?], the rational being that it is possible to calibrate cameras from the data available during use rather than having to rely on special purpose calibration data. This is an admirable endevour but all of these techniques must be evaluated on the basis of the above criteria if they are to form the basis of a reliable automatic system. Unfortunately, numerical robustness is often sacrificed for mathematical sophistication. In general many optimal solutions to the calibration problem exist [?, ?] and provided any specific method minimises the correct quantities with the right number of degrees of freedom the resulting calibration will be as good as any other formulation of the problem. The only thing to be gained from this point onward is speed of computation.

On this particular topic it only remains to be said that sometimes algorithms are requi red for problems for which no camera model exists, such as matching features in NMR images taken at two different moments in time. Although it is correct to refer to the resulting algortihms as correspondance matching they are not strictly a "true" stereo algorithm. These problems will clearly require a 2D search and this can only be achieved if the image data has photometric properties. To conclude, given that automatic camera calibration is a reality, any "true" stereo ma tching algorithm which does not make use of the epi-polar geometry will be wasting a very useful constraint on the possible solution.

3 The Principles of Area Based Algorithms

Given any two views of the same scene it can be seen that, at some image scale, a degree of similarity exists between the two views, and in general, the coarser the scale the more similar the views become. If a view is spatially quantised into ever smaller subregions, eventually as the number of features decreases, any given subregion will begin to look more similar to its corresponding subregion in the other view. Thus, by quantising a view into a number of subregions or blocks, or by changing the scale of the view in question, it is possible to apply an area based similarity metric to find the most likely correspondence between the same regions from two different views. If similarity is defined as the dot-product score of two vectors $L = (a, b)$ and $R = (c, d)$ where all bins are non-negative, then it can be said that their dot-product score will always increase if the vectors are coarsened by integrating bin contents [1]. The dot-product is given by:

$$L.R \;=\; ac + bd$$

Coarsening L and R, by bin integration, now gives two new terms $M = a + b$ and $S = c + d$ respectively, therefore:

$$M.S \;=\; ac + bd + ad + bc \;=\; L.R + ad + bc$$

Since the coarser dot-product contains extra terms, it will always be larger than the original. The use of an appropriate similarity metric is fundamental to area based stereo methods as a means for enforcing a local figural consistency constraint, surface smoothness is also usually assumed. The concept of a cross-correlation function and a search space is also introduced with the use of area based methods. Similarity metrics derived from probability density functions do theoretically offer the best solution although in practice, approximations such as Euclidean distance and dot-product metrics have been used.

4 Survey of Area Based Stereo Algorithms

4.1 Levine M.D. et al. 1973

This work forms the basis for a robot control system capable of exploring its environment [20]. The epipolar constraint is used to restrict the correspondence search to a search along a single epipolar line. The algorithm uses the classical correlation measure as shown in eqn 14.1 which is applied within a variable size window.

$$C_{X,Y,D} \;=\; \frac{\sum_y^N \sum_x^N (Left_{x,y} Right_{x+D,y} - \mu_L \mu_R)}{\sqrt{\sum_y^N \sum_x^N (Left_{x,y}^2 - \mu_L^2)}\sqrt{\sum_y^N \sum_x^N (Right_{x+D,y}^2 - \mu_R^2)}} \tag{1}$$

Where X and Y define a block location in the left image and D is the disparity parameter. The model of the world assumes that the surface consists of an approximately horizontal plane which extends to infinity. Given this well constrained model the algorithm seeks to build up horizontal contours which describe the ground plane using points called "tie-points". In order to achieve computational efficiency a coarse search of the correlation space is performed to identify candidate areas for the solution. A fine search is then performed at the most likely candidate areas. This assumes that a data model of the correlation function surface is known, thus enabling a suitably robust quantisation step to be determined for the coarse search.

4.2 Mori K. et al. 1973

Most of the earliest work in stereo vision is concerned with the processing of aerial photographs for terrain depth estimation. The work of Mori et al. [23] involves the development of a system for such problems which incorporates some basic principles. The use of a normalised dot-product score as in eqn 14.3 with the addition of a gaussian weighting is used to correlate the registered images.

The dot-product cross-correlation function is a similarity metric used in many applications where the maximum score is taken to represent the match with the highest probability of being the correct stereo match. A weighting term $w(x, y)$ consisted of a 2D gaussian kernel used to give priority to (in an ad hoc fashion) components in the centre of the calculation. In this case the dot-product score is normalised against the auto-correlation score for the respective regions in the left and right images. The use of the normalised dot-product results in the cosine of the angle between the two vectors being calculated as in eqn 14.2, the resultant measure is still image scale invariant.

$$\cos(\theta) \;=\; \frac{A.B}{|A||B|} \qquad \textit{where } A \textit{ and } B \textit{ are } N \textit{ dimensional vectors.} \tag{2}$$

The Mori et al. algorithm also demonstrates the use of high confidence estimates, variable window sizes and iterative refinement of an interpolated disparity map. The algorithm assumes that edges within the images represent surface detail and not depth discontinuities. This assumption allows regions containing edges to be treated as high confidence estimates, and from these starting points a solution is propagated. Once the solution is propagated the estimate of depth is used to reconstruct the right image. The new right image is then recorrelated with the original right image and the process is iterated to refine the solution.

4.3 Hannah M.J. 1985,1989

Hannah's work involving SRI's stereo system [12, 13] suggests the application of a number of different algorithms and incorporates the ideas of interest operators, hierarchical processing, area correlation and left-right consistency. The Moravec Interest Operator is proposed as a means of locating points within the image for which a high level of confidence about their matchability is assumed. Stereo matching by cross-correlation is applied to a subset of regions centred on peaks in the interest function. A correlation function used for this process is the normalised dot-product function as in eqn 14.3.

$$C_{X,Y,D} = \frac{\sum_y^N \sum_x^N Left_{x,y} Right_{x+D,y}}{\sqrt{\sum_y^N \sum_x^N Left_{x,y}^2}\sqrt{\sum_y^N \sum_x^N Right_{x+D,y}^2}} \tag{3}$$

Where X and Y define a block location in the left image and D is the disparity parameter. High confidence matches are evaluated first and used to steer further matching. The interest operator used by Hannah calculates the product of local intensity variance and a heuristic directional variance quantity. The use of hierarchical processing is employed by using matching at coarser resolutions to "set the context" for matches at higher image resolutions. This work is a clear example of how different methods can be applied cooperatively, although at the core of the algorithm is the reliance on luminance cross-correlation. Additionally, the hierarchical processing constraint leads to the assumption that the method assumes a surface model which is both locally and globally smooth. In summary, the algorithm seems more suited to estimating depth from aerial photographs than, for example, industrial environments.

4.4 Inria 1991

Based on the findings of Guelch [10], this work attempts to address the problem of obtaining dense depth data using region based similarity metrics and a number of other constraints [15]. The aim of the work is to obtain as dense a depth map as possible with the initial correlation phase, and to ensure that "no answer" is returned if the match fails to pass a left-right consistency test. A full depth map is then obtained by interpolating the already dense data. This work also incorporates implementation issues by developing the algorithm to run on parallel hardware with a simplistic control flow. The concept of locating regions of the image which are considered worth matching is also rejected, instead an attempt is made to match the whole image with "invalid" matches being rejected afterwards in the validation phase. The work discusses two correlation functions, a Euclidean distance based metric and a Euclidean distance metric normalised by the mean va! ! ! lue of the region. It is suggested that the extra normalisation provides invariance to linear transformations of the image grey-levels. The normalised form of the Euclidean distance metric used is shown in eqn 14.4.

$$C_{X,Y,D} = \frac{\sum_y^N \sum_x^N ((Left_{x,y} - \mu_L) - (Right_{x+D,y} - \mu_R))^2}{\sqrt{\sum_y^N \sum_x^N (Left_{x,y} - \mu_L)^2}\sqrt{\sum_y^N \sum_x^N (Right_{x+D,y} - \mu_R)^2}} \tag{4}$$

Where X and Y define a block location in the left image and D is the disparity parameter. The work discusses alternatives for the correlation validation stage and rejects the ideas of thresholding the normalised correlation score and examining the peak in the correlation surface. The left-right consistency constraint is used whereby the hypotheses obtained by matching from each image independently must reinforce each other.

4.5 Otto G.P. and Chau T.K.W. 1989

In contrast to the area based approaches discussed so far, the approach used here is to grow the solution outwards from assumed correct seed points [26]. The algorithm is based on Gruens algorithm which uses an approximate planar model for the world. This model only allows affine deformation between small regions within the two views, thus assuming that perspective effects can be ignored at a regional level. The best match is found by optimising a set of parameters which minimises the least squares error between the two views in terms of the luminance levels. The Otto et al. algorithm uses Gruens algorithm by iteratively applying it to some initial solutions within the images, and then using the results to propagate the solution to neighbouring regions. Although a model for superimposed noise on top of the image luminance is allowed for, this class of algorithm is clearly only suitable when the photometric invariance between views can be relied upon.

4.6 Okutomi M. and Kanade T. 1992

This work [25] addresses the issues of window size selection and presents a statistically sound technique which minimises the uncertainty in the disparity estimate at each pixel of the depth map. The authors make the observation that larger matching windows provide better disambiguational ability and less accuracy, and attempt to address this problem by optimally selecting a window size in a dynamic fashion as the matching process proceeds. This demonstrates the use of a principled technique for determining an otherwise arbitrary parameter. Although, in practice, it would seem that either a hierarchical matching technique or a deformable surface model would have the same effect.

5 Summary of Area Based Approaches

In their most simplistic form, area based approaches involve subdividing the whole view into subregions and applying a photometric similarity measure to all regions [20]. The aim of this type of algorithm is usually to return a dense depth map, whereby a depth estimate is made at every pixel within the scene. This approach is generally only applicable to the class of stereo problems where the following criteria are satisfied: The lighting source must ideally be a point source at infinity; the surfaces in the scene should ideally be Lambertian; the amount of figural dissimilarity or distortion between the views is small. As is stated in the aims of this work in section **??**, in many situations where stereo has applications, idealised environments and light sources cannot be assumed. The criteria developed for assessing the location accuracy of stereo algorithms will reveal that local errors can become large in the absence of a rotational distortion model.

The problem with using a single depth estimate to describe the depth over a finite subregion of the image is that this quantisation introduces location errors as described above. Area based approaches which attempt to address this problem are often referred to as window shaping techniques [26]. Whilst the field of application for these techniques is still for estimating dense depth maps, the location accuracy which can be achieved is not limited by block quantisation effects. However, the reliance on luminance consistency is, if anything, more important to the success of this class of relaxation algorithms.

In general, existing area based stereo techniques provide data which is locally inaccurate due to the lack of figural deformation invariance. Additionally, their lack of photometric invariance restricts their use to problems where grey-level consistency exists between views, however, if these constraints are satisfied they do deliver a more dense depth map than feature based approaches.

6 The Principles of Feature Based Stereo

Many areas of computer vision, such as stereo, object recognition and object tracking exploit a feature based approach. The definition of a feature is arbitrary and the only real generalisation which can be made is that a feature must be in some sense a useful parameterisation of the image. In general, useful features must have the following properties: uniqueness, repeatability and physical meaning. In the context of stereo the aim of these properties is to provide unambiguous matches with a degree of noise immunity. Since stereo vision involves extracting three dimensional data (3D data) from the scene, the features which are useful in the stereo sense are features which describe the underlying 3D structure of the scene. In the main, and particularly in man-made environments, the underlying 3D structure is described by the edges and by edge intersections (one definition of a corner). For this reason much of the feature based stereo work involves the extraction of edgels and ! ! ! corners. In some cases the extracted edgels, which are obtained using something similar to Canny [5], are linked into high-level data objects called edge-strings, stereo matching then proceeds at the edge-string level.

7 Survey of Feature Based Stereo Algorithms

7.1 Barnard S.T. and Thompson W.B. 1980

The algorithm discussed in this work [4] initially locates “interesting” points as match primitives independently in each view, where interesting is defined as those points which have a high value of variance in all four surrounding directions. In practice, an operator such as the Moravec operator [22] is suggested as being a suitable candidate feature detector, although more robust techniques for locating interesting points do exist [14]. The choice of feature is such as to provide match primitives which (a) are distinct from neighbouring points, for uniqueness and

(b) remain consistent between views, for identification purposes. Having detected matchable primitives, a mesh type data structure is created which describes potential correspondences. The initial likelihood estimate for the hypothesised correspondence is obtained by calculating a local similarity metric between the features in question. An iterative relaxation technique! ! ! is then applied to the initial likelihoods by applying a local continuity constraint. This technique provides a sparse depth map and is dependent on the repeatability of the feature detector, however, the technique is applicable to both stereo and temporal matching.

7.2 Pollard S.B. 1985

In his PhD. Thesis [27] Pollard discusses an edge based stereo algorithm called PMF which relies strongly on local support provided by a disparity gradient constraint. The algorithm detects edge primitives (edgels) in the left and right images and in the first instance defines compatible matches as those which are consistent with epipolar geometry. For each potential match a "goodness" value $C_{pp'}$ is calculated. It is suggested that $C_{pp'}$ could be calculated from edgel parameters such as edge orientation and contrast. It is also suggested that $C_{pp'}$ could be taken as the contrast of the weaker primitive in order to bias against weak edges. The statistical alternative for $C_{pp'}$ might be to directly use a characteristic PDF of the edgel parameters to define match compatibility [2], although, in practice, a usable PDF may not be obtainable. Local consistency is enforced by calculating a match strength quantity defined in eqn 14.6, which applies a disparity gradient constraint in a circular neighbourhood centred on the match hypothesis. The use of the disparity gradient is defined in cyclopean space by eqn 14.5.

$$DG = \frac{h_r - h_l}{\sqrt{h_c^2 + v_c^2}} \tag{5}$$

Where h_l and h_r are the left and right horizontal differences between a pair of points when imaged in a parallel camera geometry and h_c and v_c are the horizontal and vertical differences between the same pair of points when imaged in a virtual cyclopean space.

$$MS_{pp'} = \sum_{all\ i} max\ \frac{C_{ij'} DG(M_{pp'}, M_{ij'})}{S(M_{pp'}, M_{ij'})} \qquad M_{ij} \in N(M_{pp'}) \tag{6}$$

Where $N(M_{pp'})$ is the set of matches in the cyclopean neighbourhood of match $M_{pp'}$, and $S(M_{pp'}, M_{ij'})$ is the cyclopean separation between match $M_{pp'}$ and match $M_{ij'}$. The algorithm also uses an iterative relaxation strategy to select the final solution set from the possible hypotheses by applying uniqueness and ordering. The surface model adopted by this version of PMF allows deformation between views but does not attempt to treat edge-strings as single entities.

A later algorithm developed after the original PMF does treat edge-strings as single matchable entities. An edge-string is obtained by firstly running an edgel extraction algorithm (such as Canny [5]) followed by an edgel linking phase. The results of this process are sets of list type data structures called edge-strings. Figural support is then achieved by accumulating a match along the length of the string to give a string match score. This type of approach is not only extremely flexible but provides perhaps the most appropriate model for stereo matching in industrial environments. The model adopted by edge-string based matching is simply that the world consists of sets of edges which are subject to usually minor figural deformations between stereo views.

7.3 Ohta Y. and Kanade T. 1985

This work uses dynamic programming; firstly at the image raster level and then on the whole image to find a global solution. Image rectification is used to align epipolar lines with image rasters [24]. The technique uses edgel primitives as the basic match delimiters and computes a matching cost for a given edge delimited interval based on the combined variance measure of the luminance values within the two intervals. Given that this cost function is evaluated for all potential interval correspondences in all of the scanlines, the problem then becomes one of picking the best set of interval matches which enforced both horizontal and vertical consistency. Dynamic programming provides an optimal solution for this global consistency problem. This work defines dynamic programming as follows.

> "Dynamic programming solves an N-stage decision process as N single-stage processes. This reduces the computational complexity to the logarithm of the original combinatorial one. In order to apply dynamic programming, however, the original decision process must satisfy the following two requirements. First, the decision stages must be ordered so that all of the stages whose results are needed at a

given stage have been processed before then. Second, the decision process should be Markovian: that is, at any stage the behaviour of the process depends solely on the current state and does not depend on the previous history."

The selection of an appropriate cost function for the nodes in the dynamic programming array is perhaps the only difficulty with this algorithm. The measure used in this work assumes grey-level consistency between views, this is deemed appropriate given that the application area is the estimation of depth from aerial photographs.

7.4 Baker H.H. 1982

Baker [2] adopts the same model as that used by Ohta et el. [24], in that it is assumed that the scene consists of matchable edgels which are aligned with image rasters after image rectification. Baker uses many probabilistic measures which are combined into an overall likelihood of a given match hypothesis, this raw likelihood data is then fed into a dynamic programming algorithm which attempts to solve for a given raster. Global contour constraints are then applied to enforce inter-raster edge continuity, this differs from the work of Ohta et el., which additionally uses dynamic programming for inter-raster consistency. The raw likelihood data is based on the static edge properties such as the contrast and orientation of the edges being matched. The probability values are obtained by integrating a unit portion of an observed characteristic PDF. For example, to calculate the probability that an edge in the left image, with edge contrast CL_i, correspo! ! ! nds to an edge in the right image, with edge contrast CR_j, eqn 14.7 is evaluated.

$$P_{ij}(x) \;=\; \int_{x-0.5}^{x+0.5} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\left(\frac{x-\mu}{\sigma}\right)^2\right] \qquad where\ x \;=\; CL_i - CR_j \tag{7}$$

Probabilities for interval intensity correspondence are also calculated. The work of Baker is both comprehensive and statistically well principled, as well as being highly applicable to the stereo problem in man-made environments.

7.5 Mclauchlan P.F. 1990

In general feature based algorithms are assumed to be more appropriate when the quantity of features is relatively small. However, Mclaughlan [21] shows that extracted edgels can be used for stereo matching in situations where a large amount of texture either exists naturally, or has been imposed artificially. This algorithm firstly grids up the left image into rectangular blocks and then by constructing a disparity histogram from the set of hypothetical matches derived from all edgels within the block. The edge match criteria are based on edge orientation, contrast and a disparity gradient limit. Since the quantity of edgels within a block is necessarily large, peaks in the disparity histogram contain sufficient data to be statistically representative of the most likely disparity values for the block. Having established the most likely disparity value for the block, the data around the histogram peak is used to hypothesise a plane using a Hough transform techn! ! ! ique. After this stage extensive smoothing processes are used to enforce a local surface smoothness constraint and a global surface continuity constraint. This algorithm demonstrates the use of voting techniques and the effective use of edgel data in highly textured scenes.

8 Summary of Feature Based Approaches

In many respects feature based algorithms are established as the most robust way to implement stereo vision algorithms for the class of problems defined above as being industrial stereo problems. The advantages offered by using features are that feature based representations contain desirable statistical properties and provide algorithmic flexibility to the programmer. The flexibility being that algorithmic constraints can be applied explicitly to the data structures rather than implicitly as with area based correlation techniques. In particular the use of edge-string based representations leads to algorithms which are as locally accurate as the precision to which the edges can be extracted.

Summarising, in comparison to area based stereo algorithms, which attempt to provide dense depth data, edge-string based stereo algorithms provide sparser depth data which is locally more accurate and globally more reliable for the following reasons. Firstly, edge-string based algorithms do not use a region based planarity model to describe the world, they instead model the world as consisting of linked edges, a more flexible model for "difficult" stereo problems where the world model must incorporate some figural deformation invariance. Secondly, edge-string based algorithms exploit the properties of edge based data, which can be extracted reliably from the scene in a way which is immune to noise and invariant to luminance variations between views caused by non-ideal lighting. Where ideal lighting is something like a point source at infinity.

References

[1] Ashbrook A.P. "Pairwise Geometric Histograms", ESG MPhil. to PhD. Transfer Report, Department of EEE, University of Sheffield, Sheffield, 1993.

[2] Baker H.H. "Depth from Edge and Intensity Based Stereo", Stanford AI Lab., Stanford University, 1982.

[3] Ballard D.H., Brown C.M., "Computer Vision", Prentice-Hall Inc., 1982.

[4] Barnard S.T., Thompson W.B., "Disparity Analysis of Images", IEEE Trans. of PAMI, vol. PAMI-2, 4, 1980.

[5] Canny J.F., "A Computational Approach to Edge Detection", IEEE Trans. of PAMI vol. PAMI-8, 6, 679-698, 1986.

[6] Chevrel M., Courtis M. and Weill G., "The SPOT Satellite Remote Sensing Mission", Photogrammetric Eng. Remote Sensing vol. 47, 8, 1163-1171, 1981.

[7] Courtney P. and Thacker N., "Hardware Design for Real-time Image Rectification", AIVRU, University of Sheffield, Internal Report, 1991.

[8] Evans S., "The DIP Chip", PhD. Thesis, Dept. EEE, University of Sheffield, 1994.

[9] Goodenough J., Shelley A.J., Seed N.L., "HART, A Heterogenous Architecture for Real-time Prototyping, Development and Implementation of Machine Vision Applications", International Conference on Image Processing (ICIP), Austin, Texas. November 13-16, 1994.

[10] Guelch E., "Results of Test on Image Matching of ISPRS WG III/4", Institute of Photogrammetry, University of Stuttgart, 1988.

[11] Hannah M.J., "Bootstrap Stereo", Proc. Image Understanding Workshop, 1980.

[12] Hannah M.J., "SRI's Baseline Stereo System", Proc. of DARPA Image Understanding Workshop, 149-155, 1985.

[13] Hannah M.J., "A System for Digital Stereo Image Matching". Photogrammatic Engineering and Remote Sensing 1765-1770, 1989.

[14] Harris C. and Stephens M. "A Combined Corner and Edge Detector". Proc. 4th Alvey Vision Conference 147-151, 1988

[15] Inria Research Labs. France., "A Parallel Stereo Algorithm that Produces Dense Depth Maps and Preserves Image Features". Research Report No. 1369, 1991.

[16] Peter Burt, Bela Julesz, "Modifications of the Classical notion of Panum's Fusional Area", Perception, vol. 9 671-682, 1980.

[17] Lane R.A., Thacker N.A., Seed N.L., "Stretch-Correlation as a Real-Time Alternative to Feature Based Stereo Matching Algorithms", Image and Vision Computing Journal, vol. 12 No. 4, May 1994.

[18] Lane R.A., Thacker N.A., Seed N.L., "A Correlation Chip for Stereo Vision", British Machine Vision Conference, 1994.

[19] Lane R.A., Thacker N.A., Seed N.L., Ivey P.A., "A Stereo Vision Processor", Proc. of IEEE Custom Integrated Circuits Conference, 1995.

[20] Levine M.D., O'Handley D.A., Yagi G.M. "Computer Determination of Depth Maps". Computer Graphic and Images Processing vol. 2, 131-150, 1973

[21] McLauchlan P.F. "Recovery of Textured Surfaces using Stereo Vision", PhD Thesis, AIVRU, University of Sheffield, 1990.

[22] Moravec H.P., "Towards Automatic Visual Obstable Avoidance", Proc. 5th Int. Joint Conf. Artificial Intell., Cambridge, MA, p.584, 1977.

[23] Mori K., Kidode M. and Asada H., "An Iterative Prediction and Correction Method for Automatic Stereo-comparison". Computer Graphics and Image Processing 2, 393-401, 1973.

[24] Ohta Y. and Kanade T. "Stereo by Intra- and Inter-Scanline Search Using Dynamic Programming", IEEE Trans. on Pattern Analysis and Machine Intelligence", vol. PAMI-7, 2, 1985.

[25] Okutomi M. and Kanade T., "A Locally Adaptive Window for Signal Matching". Intl. Journal of Computer Vision 7:2, 143-162, 1992.

[26] Otto G.P., Chau T.K.W., "A "Region Growing"Algorithm for Matching of Terrain Images". Proc. 4th Alvey Vision Conference 123-128, 1988.

[27] Pollard S.B., "Identifying Correspondences in Binocular Stereo", PhD Thesis, AIVRU, University of Sheffield, 1985.

[28] Porill J., Pollard S.B., Pridmore T., Bowen J., Mayhew J.E.W., Frisby J.P., "TINA: A 3D Vision System for Pick and Place" AIVRU, 3D Model Recognition from Stereoscopic Cues, MIT Press, ISBN 0-262-13243-5, 1991.

[29] Rygol M., Pollard S.B., Brown C.R., "A Multiprocessor 3D Vision System for Pick and Place", AIVRU, 3D Model Recognition from Stereoscopic Cues, MIT Press, ISBN 0-262-13243-5, 1991.

[30] Sailesh K Rao et al., AT&T Bell, Labs. "A Real- Time P*64/MPEG Video Encoder Chip". IEEE International Solid-State Circuits Conf. 32-33, 1993.

[31] Shelley A., "HART, Pixel Port Interconnection Specifications", ESG Internal Reference Document, Dept. EEE, University of Sheffield, 1995.

[32] Thacker N.A., Mayhew J.E.W., "Stereo Camera Calibration from Arbitrary Stereo Images", 1990.

[33] Thacker N.A., Ivey P.A., "An Assessment of Image Processing and Computer Vision Algorithms Suitable for VLSI Implementations", ESG, Dept. EEE, University of Sheffield, Report 93/3, 1993.

[34] Tsai R.Y., "An efficient and Accurate Camera Calibration Technique for 3D Machine Vision", Proc. IEEE Computer Vision and Pattern Recognition, 1987.