

Evaluation of Modern Dynamic Programming Algorithms for Real-time Active Stereo Systems

J-D. Nahmias
d.nahmias@cs.ucl.ac.uk

A. Steed
a.steed@cs.ucl.ac.uk

B. Buxton
b.buxton@cs.ucl.ac.uk

Department of Computer Science, University College London
Gower Street, London WC1E 6BT, UK

ABSTRACT

Stereo Vision has been an active research field that has produced a variety of different algorithms. Unfortunately most of the algorithms that produce superior results rely on non-linear optimization techniques that are very computationally expensive, and therefore not feasible to use for real-time applications such as tele-immersion. This paper will examine a number of real-time stereo algorithms based on dynamic programming (DP) used in conjunction with structured light in order to improve the quality and facilitate the correspondence search. We will examine some of the early DP algorithms as well as the more recent work produced by [Criminisi et al] for the purpose of Gaze manipulation for teleconferencing in the context of 3d reconstruction. We present adaptation of [Birchfield et al] DP algorithm to work with structured light. Additionally we look at spatial-temporal support region for computing matching costs.

Keywords: Dynamic Programming, Stereo, Tele-immersion, Real-Time

1. INTRODUCTION

Tele-immersion relies on creating the illusion that its users are sharing a collaborative virtual space. This is a difficult task to achieve but, with recent advances in the fields of real-time 3d graphics, networking and computer vision, this has become a feasible goal and very promising results are starting to surface [Mull04]. To maintain the illusion of tele-immersion, users should perceive each other in 3d. This is not limited to viewing each other in stereo since motion parallax should be preserved (i.e. when a user moves the display should update itself appropriately rendering images from the new point of view).

Ideally, one would like to capture the 3d information in real-time, thereby removing the requirement to model deformations and not compromising behavioral realism while maintaining an accurate structural realism. Stereo vision has been a very active field of research for a few decades and has produced good results as well as a huge wealth of

literature on the subject. For an excellent overview of modern stereo reconstruction techniques readers are referred to [Scha02].

This paper will focus on examining Dynamic Programming algorithms with structured light for stereo reconstruction owing to their relatively low computational cost when compared to other stereo reconstruction algorithms based on non-linear optimization algorithms. The traditional DP algorithms [Cox96] will be presented as well as more recent variations. Their relative performance will be examined in a structured and unstructured light context. [Zhang03] proposed a framework for space-time stereo that utilizes temporal coherence in-order to further improve stereo reconstruction with structured light. The algorithms presented in this paper will similarly be extended into the space-time domain.

2. DP STEREO FORMULATION

This section will cover the standard DP stereo formulation [Cox96] as well the variations due to [Birch98] and [Crim03].

Traditional DP Stereo

Given a pair of rectified images $I_l(x)$ and $I_r(y)$ representing the left and right images at the x th and y th pixels respectively for a given scanline, it can be shown from [Forsyth02] that the depth of a given pixel is inversely proportional to its disparity $(x - y)$. The problem is therefore one of

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Uj qtv'rcrgt'rtqeggf lpi u"KDP": 2/; 25322/; /7
WSCG'2005, January 31-February 4, 2005
Plzen, Czech Republic.
Copyright UNION Agency – Science Press

correspondence. Using the *uniqueness* and *monotonic ordering* constraint DP algorithms will solve the disparity by minimizing a cumulative cost function $C(l, r)$ defined as follows:

$$C(l, r) = \min \begin{cases} C(l-1, r) + OccCost \\ C(l-1, r-1) + M(l, r) \\ C(l, r-1) + OccCost \end{cases}$$

Here, $OccCost$ is a parameter of the system and defines a penalty for occlusions and $M(l, r)$ is a cost function that defines the dissimilarity between two pixels l and r of the left and right scanline respectively. It is quite common for $M(l, r)$ to be the Sum of Squared Difference or SSD defined as:

$$M(x, y) = \sum (I_l(x) - I_r(y))^2$$

The recurrence in $C(l, r)$ defines the allowed moves in the forward pass of the DP algorithm, namely: one horizontal occluded move, one diagonal matched move, and one vertical occluded move. After initialization of the cost matrix the DP algorithm iterates through each cell within the constraint network calculating $C(l, r)$ and storing a backwards link to the previous cell containing the minimum cost. Once the cost matrix has been calculated the second stage of the algorithm is a backwards pass that follows all the stored links to produce the minimum cost path and therefore the disparity for that scanline. This is repeated for each scanline in the pair of images and a disparity map is produced.

Birchfield's DP Algorithm

The [Birch98] algorithm differs from the traditional DP algorithm in a few key ways, summarized as follows. Firstly, the cumulative cost function and the dissimilarity measure are changed and defined as follows:

$$C(l, r) = \min \begin{cases} C(l-1, r) + OccCost \\ C(l-1, r-1) + M(l, r) + MatchReward \\ C(l, r-1) + OccCost \end{cases}$$

where:

$$M(x_i, y_i) = \max \{0, I_L(x_i) - I_{\max}, I_{\min} - I_L(x_i)\}$$

$$I_{\min} = \min(I_R^-, I_R^+, I_R(y_i)), I_{\max} = \max(I_R^-, I_R^+, I_R(y_i))$$

$$I_R^- = \frac{1}{2}(I_R(y_i) + I_R(y_i - 1)), I_R^+ = \frac{1}{2}(I_R(y_i) + I_R(y_i + 1))$$

The dissimilarity function $M(x, y)$ measures how well the intensity at x fits the linearly interpolated region around y .

Another change is the addition of a constraint that intensity variation accompanies depth discontinuities. An intensity variation is defined as any set of three pixels whose min and max levels vary more than

four. This threshold is very low and is intended to prevent the algorithm from making poor choices in regions of the image that do not contain much information. It also specifies on which side the depth discontinuity must lie with respect to the intensity variation and also requires occlusions to be accompanied by the intensity variation on the appropriate side.

The cost matrix is also computed in a more efficient manner by computing the minimal cumulative cost of reaching neighbouring cells through the particular cell currently being evaluated. The computational cost is equivalent to that of the traditional DP algorithm [Cox96]. However, it permits the algorithm to prune the cost matrix and further reduce the number of cells that need to be evaluated and thereby reducing the runtime to $O(n\Delta \log \Delta)$.

Once the cost matrix has been calculated, the initial estimates of the disparities are further refined by post processing steps. Firstly outliers are removed by a mode filter. Then the disparities regions are grown in the x and y axis based on reliability, the stopping criteria is intensity variation. This paper will show in section 4 that these post-processing steps can cause problems when used in conjunction with structured light.

Criminisi's DP Algorithm

In [Crim03] a new DP algorithm is proposed with the motivation of creating a depth map in order be used in conjunction with an IBR technique that morphs two images to create a new image from a different view point. This paper evaluates this algorithm from the point of view of 3d reconstruction. The algorithm uses a three-plane graph, a left occluded plane L, a matched plane M and a right occluded plane R (see Figure 1).

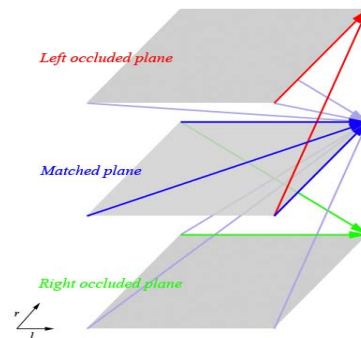


Figure 1 Adapted from [Crim03] showing the 13 allowed moves of the algorithm

This model allows a total of thirteen moves in the DP and has the advantage of allowing a much finer grain control of penalty costs. The following defines the cumulative cost functions $C_L(l,r), C_M(l,r), C_R(l,r)$ for each plane L, M and R respectively:

$$C_L(l,r) = \min \begin{cases} C_L(l,r-1) + \alpha \\ C_M(l,r-1) + \beta \end{cases}$$

$$C_M(l,r) = M(l,r) + \min \begin{cases} C_M(l-1,r) \\ C_L(l-1,r) + \beta \\ C_R(l-1,r) + \beta \\ C_M(l,r-1) \\ C_L(l,r-1) + \beta \\ C_R(l,r-1) + \beta \\ C_M(l-1,r-1) \\ C_L(l-1,r-1) + \beta \\ C_R(l-1,r-1) + \beta \end{cases}$$

$$C_R(l,r) = \min \begin{cases} C_R(l-1,r) + \alpha \\ C_M(l-1,r) + \beta \end{cases}$$

Here, α is the cost of moving within an occluded plane and β the cost of making a transition between planes. $M(l,r)$ is a windowed normalized cross-correlation: $M(l,r) = (1 - M'(l,r)) / 2$ where

$$M'(l,r) = \frac{\sum (I_L - \bar{I}_L)(I_R - \bar{I}_R)}{\sqrt{\sum (I_L - \bar{I}_L)^2 \sum (I_R - \bar{I}_R)^2}}$$

In [Crim03] the dissimilarity matrices are stacked across all the scanlines and Gaussian smoothed with a kernel orthogonal to both left and right scanlines.

3. IMPLEMENTATIONS

This section will cover the various implementations as well as certain variations that had to be made in order for the algorithms to work with structured light. A possible framework for space-time stereo using structured light described in [Zhang03] motivated the extension of the implemented algorithms into the space-time domain. However the space-time support windows were not sheared and skewed as described in [Zhang03]. The results of these various implementations will be presented in section 4.

The [OpenCV] implementation was used for the [Birch98] DP algorithm and modified while the [Crim03] DP algorithm was implemented from scratch without using Gaussian smoothing of the dissimilarity matrix. Support for SSD dissimilarity measure over spatial and temporal windows was added to the [OpenCV] implementation. The structured light patterns broke some of the assumptions made in the post-processing steps of

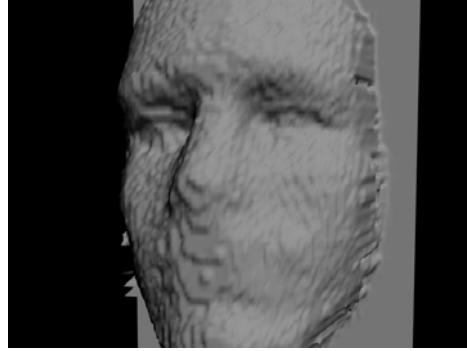


Figure 2 3d Reconstruction based on obtained disparity map using [Crim03] algorithm and space-time window as well as regular triangulation

[Birch98] described in section 2. The structured light pattern removes intensity variations along the y-axis and therefore one of stopping criteria for the region growing of the disparities along the y axis is violated. These were subsequently removed with the one exception of the mode filter.

The structured light pattern was generated as described in [Zhang03] using a 16bit Reflected Gray Code [Bit76] that is subsequently shuffled to produce high frequency change in both the spatial and temporal domains. The patterns are then smoothed and projected using an Infocus DLP projector at a resolution of 800x600 and are cycled through at the same frame-rate as the cameras performing capture.

The images were captured using two Balser A602fc cameras at a resolution of 640x480 with a quantization depth of 8bits. The camera calibration and image rectification was performed using the [Boug01] Matlab toolkit.

Finally the best disparity map was used to create 3d surface using two different triangulation algorithms. Figure 2 shows the 3d surface created using a regular triangulation.

4. EXPERIMENTS AND RESULTS

A polystyrene head was used for the experiments. The [Birch98] and [Crim03] algorithm were tested against the following dataset

- A pair of naturally lit images using spatial window.
- A pair of images lit with a structured light pattern using a spatial window
- A set of images lit with a time varying structured light pattern using a symmetrical spatio-temporal window of size 5x5x8.

The [Birch98] algorithm was tested using both SSD and interpolated cost functions, while the [Crim03] was tested using the cross-correlation cost function.

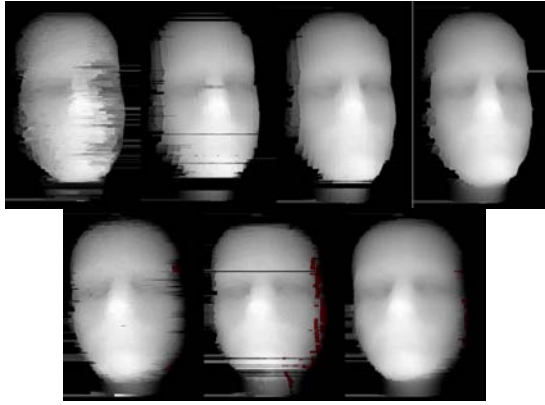


Figure 3 top left to right: [Birch98] algorithm, without light; with light; with light SSD 5x5; with light and a SSD 5x5x8 spatio-temporal, bottom left to right [Crim03], without light, with light 5x5, with light 5x5x8

All results were evaluated qualitatively by examining the disparity maps. However most of the differences in quality are quite evident on visual inspection of the disparity maps.

Figure 3 illustrates all the results. Using structure light improves the algorithm significantly, and, again, using a spatio-temporal window improves it even further. The results become interesting when comparing the [Birch98] algorithm against the [Crim03] algorithm under different conditions. Without structured light, the [Crim03] algorithm produces vastly superior results in comparison to the [Birch98] algorithm. However, when structured light is used with a spatial window the differences become much less pronounced. When comparing the performances of the algorithms with spatio-temporal windows the differences become even more subtle. Under these conditions, the performance criteria need to be clearly defined and some quantitative measures need to be applied.

From these results, one can clearly state that using the given structured light pattern improves each algorithm provided. One can make a similar statement with regards to using a spatio-temporal window with the structured light. However because the spatio-temporal window is not sheared or skewed it implicitly assumes a static scene. The statement that using a symmetric spatio-temporal window leads to an improved performance thus cannot, without further investigation, be made with regards to dynamic scenes.

5. CONCLUSION & FUTURE WORK

This paper has shown that structured light can be a powerful tool for the improvement of the DP algorithms described in Section 2. The paper has also highlighted some of the benefits in using temporal information when evaluating dissimilarities between

pixels for static scenes. However, it has also left some questions unanswered, such as whether the benefits of the [Crim03] algorithm in the unstructured case are due to the fact it uses a normalized cross-correlation dissimilarity function or may they be mainly attributed to its three plane DP graph approach. Further work still needs to be carried out to evaluate these algorithms with dynamic scenes and properly investigate whether or not the [Crim03] algorithm is superior to the modified [Birch98] algorithm when used with structured light and with a spatio-temporal SSD dissimilarity function.

ACKNOWLEDGEMENTS

The first author was supported by funding associated with the EPSRC funded Interdisciplinary Research Collaboration (IRC) project EQUATOR (EPSRC Grant GR/N15986/01)

REFERENCES

- [Bit76] J. R. Bitner, G. Ehrlich, and E. M. Reingold, "Efficient generation of the binary reflected gray code and its applications," *Communications of the ACM*, vol. 19, no. 9, pp. 517–521, 1976.
- [Birch98] S. Birchfield and C. Tomasi, Depth Discontinuities by Pixel-to-Pixel Stereo, *Proceedings of the Sixth IEEE International Conference on Computer Vision (ICCV), Mumbai, India, pages 1073-1080, January 1998*
- [Boug01] J.-Y. Bouguet. Camera Calibration Toolbox for Matlab. <http://www.vision.caltech.edu/bouguetj/calibdoc/index.html>, 2001.
- [Cox96] I.J. Cox, S.L. Hingorani, and S.B. Rao. A maximum likelihood stereo algorithm. *Computer vision and image understanding*, 63(3):542–567, 1996.
- [Crim03] A. Criminisi, J. Shotton, A. Blake, P.H.S. Torr. Gaze-Manipulation for One-to-one Teleconferencing. *Int. Conf. on Computer Vision, ICCV, Nice, France, 2003*
- [Forsyth02] D. Forsyth, J. Ponce: *Computer Vision: A Modern Approach*. Prentice Hall ISBN 0130851981
- [Mull04] J. Mulligan, N. Keshikar, X. Zampoulis, K. Daniilidis, Stereo-based Environment Scanning for Immersive Telepresence, submitted to the *IEEE Transactions on Circuits and Systems for Video Technology Volume 14, Issue 3, March 2004, pages 304-320*.
- [OpenCV] Open Computer Vision Library <http://sourceforge.net/projects/opencvlibrary/>
- [Scha02] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. on Computer Vision*, 47(1):7–42, 2002.
- [Zhang03] L. Zhang, B. Curless, and S.M. Seitz. Spacetime stereo: Shape recovery of dynamic scenes. In *Conference on Computer Vision and Pattern Recognition*, 2003.