Trinocular Stereo for Non-Parallel Configurations

Jane Mulligan and Kostas Daniilidis
GRASP Laboratory,
University of Pennsylvania,
Philadelphia, PA, USA

Abstract

The constraint of a third camera in stereo vision is a useful tool for reducing ambiguity in matching. Most of the systems using trinocular stereo to date however, have used configurations where the image planes of all three cameras are coplanar, or can be rectified to be so. In this paper we explore the computation of dense trinocular disparity maps for non-planar camera configurations which arise when cameras surround the object to be modeled. Our approach rectifies the cameras as two independent stereo pairs. We start with an exhaustive lookup scheme and then consider retaining only a list of N disparities per pixel with maximal correlation values for the right and left pairs. Experimental results and comparisons demonstrate that both methods reduce outliers over binocular stereo, and that the N–hypothesis system trades large lookup tables for somewhat lower density of valid matches.

1. Introduction

Reconstructions from a single stereo pair often have errors and extreme outliers due to ambiguity in matches along the epipolar line. For applications such as building detailed object models or creating models of humans for virtual environments, identifying and eliminating such points or patches is critical, but often difficult and expensive. One well known constraint for reducing these ambiguities is to add a third camera to verify hypothesized matches. However, most trinocular systems proposed in the literature exploit a right triangular [5, 1, 3, 4] and/or parallel [6] camera triple configuration. The close range scanning tasks we are interested in are better served by a surround configuration of cameras, which disallows the triple rectification methods which simplify these parallel camera approaches.

The trinocular epipolar constraint in stereo vision is based on the fact that for a hypothesized match $[u, v, d]$ in a pair of images, there is a unique location we can predict in the third camera image where we expect to find evidence of the same world point [2]. A hypothesis is correct if the epipolar lines in the third camera image for the original point $[u, v]$ and the hypothesized match $[u - d, v]$, intersect. For edge based systems this means checking a series of relatively sparse hypothesized matching edge points to determine which are consistent [7, 1, 3]. This correspondence and verification is simplified if the cameras are aligned or rectified in an up-down/left-right parallel right triangle configuration, where epipolar lines are made parallel to the horizontal and vertical scanlines [1]. For dense disparity calculation, Okutomi and Kanade [6] used a linear parallel configuration of cameras in their multibaseline system. They exploited a third view by summing pairwise SSD values referred to common $\frac{1}{Z}$ values.

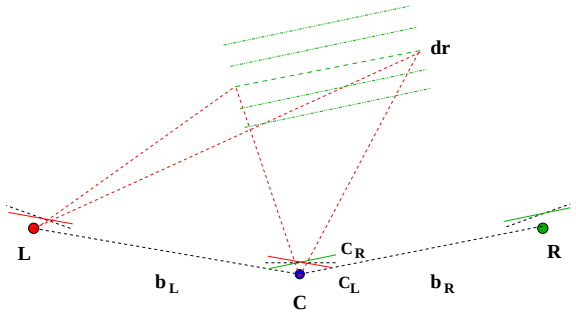

Figure 1. Trinocular camera triple.

The configurations we are interested in are similar to that depicted in Figure 1, where a sequence of cameras surrounds an object to be modeled or a user interacting with an augmented reality system. To obtain the accuracy we require for these tasks as well as facilitate merging and registering of multiple views, we use a fixed strongly calibrated camera rig. To generate dense accurate depth maps in such a scenario we implemented 2 trinocular stereo algorithms. Both treat the cameras as 2 independent pairs, left and centre and centre and right. The first algorithm combines correlation values from the two pairs by precomputing correlation images for ranges of disparity in the left camera pair, then the computed correlation for each tested $[u_R, v_R, d_R]$ is added to that precomputed for the corresponding $[u_L, v_L, d_L]$. This results in large correlation lookup tables for the left image pair. As an alternative we tested an algorithm which computes correlations over a range of disparities independently for both image pairs. It retains only the N highest correlation values for each location in each pair. Much like the edge match hypotheses

from the literature, these can be cross checked to see which are valid and have the highest correlation total. In the following sections we describe the two algorithms in detail and experimentally demonstrate and compare their results.

2. Full Calculation

We begin by independently rectifying the left and centre cameras (L and C_L) and the centre and right cameras (C_R and R), so that their epipolar lines are parallel respectively. For the right rectified camera pair every disparity d_R to be searched represents a plane with constant Z, which can be projected into the L and C_L images to compute the corresponding $[u_L, v_L, d_L]$ for each $[u_R, v_R, d_R]$. This straightforward application of the trinocular constraint is illustrated in Figure 1.

Of course for any Z-plane constructed from d_R, a range of d_L will be required to match points in the left pair. For example for the images used later, the right range $D_R = [-90, 10]$ corresponds to a left range $D_L = [-74, 67]$. Also because the two pairs are independently rectified, corresponding points in the left pair will not necessarily have $v_L = v_R$, thus all of u_L, v_L, and d_L depend on $[u_R, v_R, d_R]$. The calculation is simplified slightly by the fact that C_L and C_R are derived from the same image C and are related by the a priori rectification rotations R_{CL} and R_{CR}. We can thus precompute a lookup table of locations in C_L equivalent to those in C_R by precalculating $[u_{CL}, v_{CL}, s] = R_{CL}R_{CR}^{-1}[u_{CR}, v_{CR}, 1]^T$, for all image locations.

Our underlying matching measure is modified normalized cross correlation (MNCC) of the form:

$$\text{MNCC} = \frac{2[N\sum I_L I_R - (\sum I_L)(\sum I_R)]}{(N\sum I_L^2 - (\sum I_L)^2) + (N\sum I_R^2 - (\sum I_R)^2)}$$

for an image pair (I_L, I_R), where sums are over a correlation window of size N.

Borrowing from Okutomi and Kanade's [6] insight that we need to select matches based on minima (or maxima in the case of correlation) of the combined matching measure with respect to depth, we sum the MNCC values for corresponding $[u_R, v_R, d_R]$ and $[u_L, v_L, d_L]$ to obtain a correlation measure which now varies between -2 and $+2$.

Given the intrinsic and extrinsic camera parameters, and rectification matrices we can precalculate D_L the range of d_l generated by the plane implied by the current d_R. We calculate and store the right to left (C_L to L) correlation for the left pair, for all $d_L \in D_L$. This gives us a set c_L of $k = |D_L|$ planes of correlation values for the left centre image.

To evaluate a match at $[u_R, v_R, d_R]$, first we calculate $c_r = \text{MNCC}(C_R, R, d_R)$. For the left pair we calculate the location (u_{LL}, v_{LL}) of points on the depth plane in the left rectified image L. Using the precomputed lookup table we find the coordinates (u_{CL}, v_{CL}) and finally we can calculate the disparity $d_L = u_{LL} - u_{CL}$ for each point. Given the corresponding $[u_{CL}, v_{CL}, d_L]$ for each point in the centre right image, we can look up the correlation value c_L at the specified location in the computed left correlation planes. We can now calculate our overall correspondence by $S_{corr} = c_L + c_R$.

To summarize the algorithm:

Full Calculation:

Step 1: Precompute lookup table for C_L locations corresponding to C_R locations, and the range D_L for each d_R

Step 2: Update the left correlation lookup table c_L to include all $d_L \in D_L$ required for current d_R.

Step 3: Project world points defined by $[u_{CR}, v_{CR}, d_R]$ in the centre right image into the left image L to give (u_{LL}, v_{LL}) and lookup (u_{CL}, v_{CL}).

Step 4: Compute sum of correlations:

$$S_{corr}(u_{CR}, v_{CR}, d_R) = c_L[u_{CL}, v_{CL}, u_{LL} - u_{CL}] + \text{MNCC}(u_{CR}, v_{CR}, d_R).$$

Step 5: If $S_{corr}(u_{CR}, v_{CR}, d_R)$ is a peak in the correlation function set $D_{map}(u_{CR}, v_{CR}) = d_R$.

Step 6: Goto 2

We could also precompute lookup tables for the left disparity maps corresponding to each d_r but for images with $P = m \times n$ pixels this would expand the demands on memory to an additional $P \times (d_R^{max} - d_R^{min})$. What this system can provide is a baseline of how well our stereo reconstruction system can perform under the trinocular constraint.

3. N-MAX

We can make two observations about the large lookup tables of correlation values for the left pair: 1) many of the correlation values are likely to be low, 2) the correct match of an image point should be positively and relatively highly correlated, although it may not be at a peak in the correlation function. Based on these two insights we propose maintaining a sorted set of the highest correlation values and their disparities for the right and left image pairs.

We proceed as for usual correlation stereo, calculating the correlation function over ranges D_L and D_R for each pair respectively. Instead of maintaining only a single peak correlation value and its disparity however we maintain N disparity planes D_{map}^i and N correlation planes c_i. D_{map}^1 and c_1 will contain the disparity map and maximum correlation values corresponding to the output of the usual correlation stereo. $D_{map}^i, i > 1$ represents disparities in order of decreasing correlation value.

After performing correlation on both pairs, we examine the correlation and disparity maps. For points in each right disparity plane $D_{map,R}^i$ we predict the corresponding, $[u_L, v_L, d_L]$, then we examine the values in $D_{map,L}^i$ to determine if the predicted values also achieved a high correlation value and were retained. The valid pair $(D_{map,L}^i(u_L, v_L), D_{map,R}^j(u_R, v_R))$ with the highest sum of correlations value $S_{corr} = c_{Li}(u_L, v_L) + c_{Rj}(u_R, v_R)$ is selected.

Figure 2. Three camera views.

The algorithm proceeds as follows:

N-MAX:

Step 1: Precompute lookup table for C_L locations corresponding to C_R locations, and the range D_L corresponding to D_R

Step 2: For all $d_L \in D_L$ calculate MNCC(L, C_L, d_L), update $D_{map,L}$ and c_L.

Step 3: For all $d_R \in D_R$ calculate MNCC(C_R, R, d_R), update $D_{map,R}$ and c_R.

Step 4: For $j = 1, N$

Project world points defined by $[u_{CR}, v_{CR}, d_R] \in D^j_{map,R}$ into the left image L to give (u_{LL}, v_{LL}) and lookup (u_{CL}, v_{CL}).

For $i = 1, N$

$$\text{if} \quad |D^i_{map,L}(u_{LL}, v_{LL}) - (u_{LL} - u_{CL})| < 1$$
$$\wedge$$
$$S_{corr}(u_{CR}, v_{CR}) < c_{Li}(u_{LL}, v_{LL}) + c_{Rj}(u_{CR}, v_{CR})$$

$$D_{map,T}(u_{CR}, v_{CR}) = D^j_{map,R}(u_{CR}, v_{CR})$$
$$S_{corr}(u_{CR}, v_{CR}) = c_{Li}(u_{LL}, v_{LL}) + c_{Rj}(u_{CR}, v_{CR})$$

The N-MAX algorithm has the advantage of reducing the number of lookup tables to $2N$, but maintaining the sorted list of hypothesized correlation and disparity values will cost as much as $P \times N \times (d^{max} - d^{min})$ comparisons or $2P \times N \times (d^{max} - d^{min})$ updates. Further the cross checking of all hypotheses will cost N^2P subtractions and additions.

4. Experiments

Figure 2 shows a stereo triple used in our experiments. The subject is approximately at the vergence point of the three cameras, 80-90 cm from the image centres. We have reconstructed the scene using three methods for comparison. First we used a basic MNCC 2 camera correlation with forbidden zone constraint on the centre and right images (Figure 3). This allows us to evaluate the benefit of adding the third camera. Second we ran the full calculation version of trinocular stereo on the image triple (Figure 4), and finally we evaluated the N-MAX method (Figure 5) with varying values of N. Obviously in the limit for large N, we will retain the full set of left and right correlation maps and thus matches available to N-MAX are equivalent to the full calculation. For $N = 1$, N-MAX retains points where 2 independent stereo correlations agree on depth, and will tend to be more sparse than either the case where more hypotheses are retained or the single pair stereo.

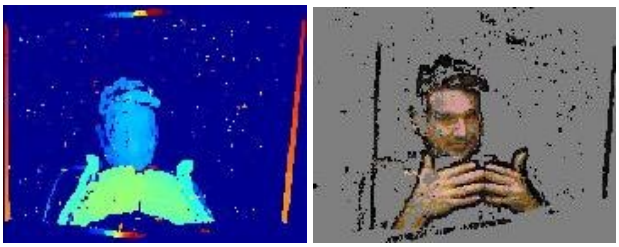

Figure 3. Disparity map and reconstructed points for 2 camera stereo.

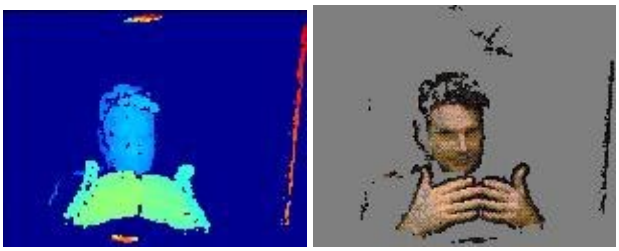

Figure 4. Disparity map and reconstructed points for full trinocular match.

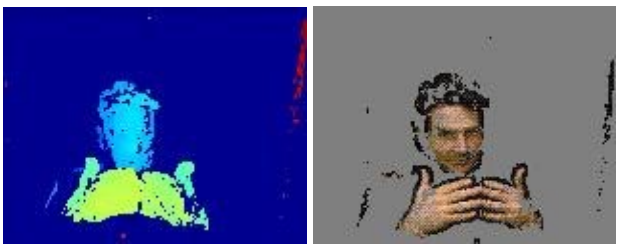

Figure 5. Disparity map and reconstructed points for multi-hypothesis, N=3.

Figures 3, 4 and 5 show the calculated disparity map and snapshot from our 3-D viewer showing a rotated view of the coloured reconstructed points. In all cases matches with low

Figure 6. Reconstructed views for binocular, full and multi-hypothesis methods rotated

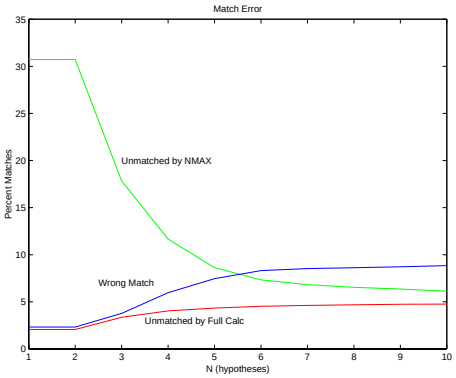

Figure 7. Match error wrt N. Percentage of valid matches for full calculation which are unmatched by N-MAX, wrongly matched or are matched by N-MAX but not the full calc.

correlation values ($< .5$ for 2 camera stereo, < 1 for trinocular) have been discarded. Figure 6 illustrates the quality of the reconstructed points for the three systems. The improvement resulting from the added constraint of a third camera is clear from the speckle of matched points in low texture areas apparent in the stereo pair reconstruction, which has been eliminated by both trinocular approaches.

The reconstructions from the full and multi-hypothesis trinocular methods look quite similar. Using the full calculation method as our best approximation of the true scene disparities, we compared the matches obtained by the multi-hypothesis system for $N = 1, 10$. Figure 7 shows a plot of percentage matching error with respect to number of hypotheses retained N. Points matched by the full calculation, but not by N-MAX drop sharply as N increases from 2 to 6. Points where N-MAX finds a different match ($|d_{NMAX} - d_{full}| > 1$) from the full calculation increase as N increases, but level off at about $N = 6$. Points matched by N-MAX but not the full calculation increase slightly, but remain fairly steady below 5%. These latter are probably the result of using the forbidden zone constraint for the full calculation, which cannot be used in the multi-hypothesis case.

5. Conclusion

Trinocular stereo reduces ambiguous matches and hence the outliers often observed in 2 camera stereo. Many of the most efficient approaches to exploiting the trinocular constraint however, involve camera configurations which can be rectified or engineered such that the image planes are parallel to the plane of the optical centres. We are interested in close range reconstruction and modeling of people and objects by a surround configuration of cameras not amenable to this type of rectification. In this paper we have explored two algorithms which exploit the trinocular constraint in these scenarios. Both rectify and perform correlation on the left and centre and centre and right camera pairs independently. The full calculation method precomputes correlation planes for the left pair, then uses the trinocular constraint to lookup the corresponding correlation value for every tested match $[u_R, v_R, d_R]$. This method requires large (up to 60 planes) lookup tables of left correspondence values. As an alternative we proposed a method which computes correlation for a range of disparities for the right and left pairs, but retains only the N highest correlation values for each location for each pair. Our experiments showed that outliers were significantly reduced over 2 camera correlation stereo. The N-MAX hypothesis method reduces outliers well, but finds fewer valid matches than the full trinocular calculation. The tradeoff in reducing the retained correlation planes to N then appears to be in losing density of valid matches, versus the space and time required to manipulate $>> N$ lookup tables.

References

[1] N. Ayache. *Artificial Vision for Mobile Robots: Stereo Vision and Multisensory Perception*. The MIT Press, Cambridge, MA, 1991.

[2] U. Dhond and J. Aggarwal. Structure from stereo – a review. *IEEE Transactions on Systems, Man and Cybernetics (SMC)*, 19(6):1489–1510, Nov/Dec 1989.

[3] O. Faugeras. *Three-Dimensional Computer Vision: A Geometric Viewpoint*. The MIT Press, Cambridge, MA, 1993.

[4] D. Murray and J. Little. Using real-time stereo vision for mobile robot navigation. *Autonomous Robots*, 2000. To Appear (http://www.cs.ubc.ca/spider/little/links/robuds.html).

[5] Y. Ohta, M. Watanabe, and K. Ikeda. Improving depth map by right-angled trinocular stereo. In *Proceedings of the 8th International Conference on Pattern Recognition (ICPR'86)*, volume I, pages 519–521, Paris, France, Oct. 1986.

[6] M. Okutomi and T. Kanade. A multiple-baseline stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 15(4):353–363, April 1993.

[7] M. Yachida, Y. Kitamura, and M. Kimachi. Trinocular vision: New approach for correspondence problem. In *Proc. 8th Intl. Conf. on Pattern Recognition*, pages 1041–1044, Paris, France, October 1986.