3D reconstruction from images taken with a coaxial camera rig

Richard Kirby\*and Ross WhitakerUniversity of Utah School of Computing, 50 S. Central Campus Drive Room 3190

SLC UT 84112

Abstract

3D reconstruction from images taken through an endoscope or borescope has numerous medical and industrial applications, but until now has not found wide acceptance due to the lack of physical space for a traditional stereo baseline. While it is possible to acquire image pairs taken along a common optical axis using a coaxial camera rig (two cameras that image along the same optical axis via a beam-splitter), performing 3D reconstruction on these images has not been possible in the center region of the images due to the very small disparity between corresponding points. This characteristic of coaxial image pairs has been called the unrecoverable point problem. We introduce a novel method to overcome the unrecoverable point problem, using a variational methods optimization algorithm to map pairs of optical flow fields from different focal length cameras in a coaxial camera rig. Instead of using pixel based correspondences our method uses the ratio of the optical flow fields at each pixel location to perform 3D reconstruction. This not only results in accurate image pair alignment but also produces accurate dense depth maps throughout the field of view. We test our method on synthetic optical flow fields and on real image sequences taken with a coaxial camera rig. We demonstrate our method's accuracy by evaluating our results against a ground-truth. Accuracy is comparable to a traditional binocular stereo camera rig, but without the need for the traditional stereo baseline and with substantially smaller occlusions.

**Keywords:** Stereo endoscope, stereo borescope, 3D reconstruction, variational methods, coaxial camera rig, depth from zooming

# INTRODUCTION

3D reconstruction from image pairs taken from two different perspectives is one of the most active areas of research in computer vision [10, 18, 20]. The most common two camera rig is the binocular stereo rig where the cameras are oriented with their optical axes parallel and separated by a baseline. 3D reconstruction from images taken with a binocular stereo rig requires finding pixel pairs (one from each camera), that are a projection of the same point in the scene [9, 26]. When the camera geometry is known, depth can be estimated using the disparity between the pixel pairs. However, pixels are finite in size, resulting in discrete disparity steps and discrete steps in the depth estimate. Thus the resolution of the depth estimate is a function of the camera baseline. The larger the baseline, the higher the resolution of the depth estimate. Larger baselines, however, create two well know problems: 1) The larger the baseline, the greater the occlusions (areas of the scene where one camera cannot see what the other camera sees) and 2) the larger the baseline, the larger the stereo camera rig.

There are computer vision applications where the traditional binocular stereo baseline is problematic, most notably in applications requiring that the camera rig be inserted into a small space, like the barrel of an endoscope/borescope or in applications where the surface being analyzed is so close to the cameras that sufficient overlap between images is impossible. Traditional binocular stereo endoscopes exist, but either the cameras are so close together that the depth resolution is low, or the instrument is too large for some applications.

One alternative to a traditional binocular stereo rig is a zero baseline camera, sometimes called depth from zooming [?] or coaxial [17] camera rig. In this type of camera rig, images are taken at two different focal lengths along the same optical axis. This creates a disparity which is a function of both the distance to the point in the scene being imaged as well as the distance that the pixel under evaluation is form the optical center of the camera. This type of camera rig produces results similar to a traditional binocular stereo rig near the edges of the images, but in the center region, the disparities are too small to produce acceptable resolution [26].

In this paper we introduce a novel automated method for finding depth in image sequences taken with a coaxial camera by using the optical flow fields. We apply the technique to both synthetic optical flow fields and real images taken with an RGB-RGB coaxial camera rig. In applications where there is sufficient motion between the camera rig and the scene (moving endoscope or borescope) and where the scene exhibits enough texture to produce optical flow, our method finds correspondences between the flow fields and uses the ratio of the flow fields at these corresponding points to estimate depth. The resulting dense depth maps are used to perform 3D reconstruction of the scene with accuracies similar to, and in certain cases, better than, techniques that align images based on image features or pixel intensities.

# RElated work

Depth from images taken at different focal lengths along a common optical axis was first proposed by Ma and Olsen [17]. Lavest et al. [16, 15] provide a proof for inferring 3D data from images taken at multiple focal lengths along a common optical axis and models a revolving object. Asada et al. [1] and Baba et al. [2] present a method for doing 3D reconstruction using blur from zoom. Gao et al. [8] present a distance measurement system for mobile robots using zooming. Most recently, Zhang and Qi [25] describe a method for 3D reconstruction from multi-focal length images using a snake-search algorithm.

The original reason researchers focused on using a single camera at different focal lengths to do 3D reconstruction was been cost. However, there are several other advantages. Ma and Olsen alluded to the fact that a depth from zoom camera exhibits substantially smaller occlusions than an equivalent binocular stereo camera rig. Additionally, there are applications where a stereo baseline is prohibitive (endoscope or bore scope) and where the known correspondence point on the optical axis is an advantage to image registration. Finally, where image registration is the ultimate objective of the application (e.g. alignment of images from two different types of sensors without attempting 3D reconstruction), a coaxial camera produces substantially smaller disparity errors in the center region than a binocular mulit-modal stereo rig.

The coaxial camera rig [11] is equivalent to simultaneous depth from zooming, but instead of changing the focal length of a single fixed camera, two cameras are arranged such that the cameras form images along the same optical axis. This is done by splitting the optical path with a beam splitter and aligning the two cameras such that their optical centers image the same point in the 3D scene. The coaxial camera rig combined with image correspondences derived from perceived motion overcomes the two main problems of depth from zooming. First, simultaneous images taken at two different focal lengths overcomes the stationary scene constraint of depth from zooming. Second, using the flow field to align image pairs overcomes the unrecoverable point problem in the center region that was described by Ma and Olsen. This later advantage is due to the depth estimate being derived from the ratio of the flow fields taken at different focal lengths as opposed to the extremely small disparities found in the center region of a coaxial camera rig.

# variational model

Referring to Figure 2, let , represent points in the image domain of the front and back cameras. Let the disparity between and such that and represent the same point in the scene. Let the focal lengths for the front camera and back cameras and the distance between the optical center of the front camera and a point in the scene correspond to , the distance being measured along the optical axis. the distance between the optical center of the two cameras. the projection of the 3D motion field onto the image planes of the front and back cameras respectively.

# Figure3

Figure 2: Coaxial camera rig geometry.

Using the projection equation to project the start point ( and end point ( of a point in the scene onto points in the image planes of each camera gives:

(1)

(2)

Where the second subscript of the points in the image plane represents the start or end of the projected motion.

Solving equations (1) and (2) for and setting them equal to each other gives:

(3)

where:

(4)



and

(5)



has a direct physical interpretation. From (5), it can be seen that if or when . Referring to Figure 3 one can see that a change in Z introduces a slight parallax () in the finishing points of the optical flow detected by the two cameras. corrects for the parallax and can also be solved for directly from the coaxial camera geometrically.

The first term in our coaxial camera variational model is an optical flow matching term:

(6)

The second term is a smoothness term:

(7)

The total energy that we want to minimize is:

(6)

where and are tuning constants.

# Figure4

Figure 3: Parallax caused by ΔZ in a coaxial camera rig.

# Numerical Solution

## Euler-Lagrange

We minimize the energy by taking the Euler-Lagrange equations for (6) and (7) with respect to z and setting to 0. The Euler Lagrange for (6) and (7) is

(8)

where prime indicates the derivative with respect to Z, is the Laplacian operator and

(9)

We reduce the problem to a 1D optimization problem by observing that the solutions lie on radial epipolar lines. The Euler-Lagrange equations (one along the radial line and the other perpendicular to the radial line) are solved using the gradient decent method.

## Initialization

We initialize the value of Z by observing that the optical flow vectors which start and end on the optical axis (e.g. or ) result in a simplified version of (3) which does not depend on :

(8)

Using , assuming ΔZ is small relative to Z and that the scene is rigid, we use the optical flow to estimate Z for all pixels in the images. For rigid scenes with no Z translation, this is identical to the optimal solution to the Euler-Lagrange equations if the optical flow fields are equivalent to the motion fields. Where there is ΔZ and/or where the scene is not rigid, this produces a good starting point for the fixed point iterations.

## Resampling to a discrete grid

The gradient descent results in a new estimate of Z at , which is offset spatially in the image domain from the previous estimate of Z by the optical flow. Since optical flow algorithms produces sub-pixel flow values, the new Z values are rarely on integer pixel locations. This requires resampling the newly estimated depth map onto an integer pixel grid to obtain the Z that corresponds to each pixel.

## Stopping Criteria

We used one of two stopping criteria depending on the quality of the flow fields and the value chosen for α. When the flow fields closely represent the motion fields and α is small (minimal Z smoothing), we use equation (7), which represents the mismatch in registration of the two flow fields, and stop when this number becomes suitably small.

Where the flow fields are noisy and not as good a representation of the motion field we need to increase α to get good results. With more substantial smoothing, the smoothing term, equation (27), can pull the Z estimate away from the correct value if γ is large and/or if many iterations are performed. In this case we stopped the iterations when the smoothing term (the second term) in equation (13) was approximately equal to, but of opposite sign to the match term (the first term) in equation (13). This later approach often results in a larger error between the flow fields vs. the first approach, but our experiments show that it results in more accurate alignment because we stop iterating before the smoothness term pulls Z too far from its correct position.

## Algorithm

* Compute and .
* Smooth and .
* Initialize Z.
* For each epipolar line:
  + Iterate
    - Update Z estimate along epipolar line by updating the previous value of Z using (8)
    - Resample Z estimate to grid
  + Has stopping criteria been met?
* Repeat for next epipolar line

## Margins

If the paper does not have the margins shown in Table 1, it will not upload properly.

Table 1. Margins and print area specifications.

**PAPER MARGINS**

A4 Letter

Top margin 2.54 cm  *(1in.)* Top margin 1.0 in.  *(2.54 cm)*

Bottom margin 4.94 cm *(1.95 in.)* Bottom margin 1.25 in. *(3.17 cm)*

Left, right margin 1.925 cm *(.76 in.)* Left, right margin .875 in.  *(2.22 cm)*

Printable area--all text, figures and footnotes:   
A4: 17.15×22.23 cm Letter : 6.75×8.75 in.

## Fonts

Table 2 shows the font sizes and highlighting in a typical manuscript. These font "styles" are contained with this sample manuscript and Section 4 below explains how to use them. Use Times Roman or another standard font to avoid font errors.

\*s.greg.jones@narelab.com; phone 1 222 555-1234; fax 1 222 555-876; narelab.com   
Table 2. Manuscript font sizes and formatting. This Microsoft Word template includes these formats as automated "styles", which can be selected in the Format menu -- Styles and Formatting. (9pt)

## Preformatted SPIE Template Styles

## *Manuscript component Description*

## SPIE paper title 16 pt. bold, center

## SPIE authors/affiliations 12 pt., center

**Keywords:** \*Keywords\* 10 pt., left justify

### SPIE ABSTRACT TITLE 11 pt. BOLD, center

SPIE abstract body text 10 pt., justify

### SPIE SECTION HEADING 11 pt. BOLD, center, whole numbers

**1.1 Heading 2, SPIE subsection heading 10 pt. bold, left justify, subsection numbers**

SPIE body text 10 pt., justify

SPIE figure caption 9 pt., justify below figure

SPIE table caption 9 pt., justify above figure

SPIE footnote text 9 pt., justify, numbered

**REFERENCE HEADING 11 pt. BOLD, center**

1. SPIE reference listing 10 pt., justify, numbered in brackets

# Formatting of manuscript components

## Title

Center the paper title at the top of the page in 16-pt. bold. Only the first word, proper nouns, and acronyms are capitalized. Keep titles brief and descriptive. Spell out acronyms unless they are widely known. Avoid starting with articles or prepositions, e.g., “The study of … ,” or, “On the ….”

## Author affiliations

The author list is in 12-pt. regular, centered. Omit titles and degrees such as Dr., Prof., Ph.D., etc. The list of affiliations follows. Each author's affiliation should be clearly noted.

## Section headings

Section headings are 11-pt. bold capitals, centered. Sections numbers have whole numbers, e.g., 1, 2, 3… Don't number the "Acknowledgements" and "References." Headings often used are: 1. Introduction, 2. Methodology, 3. Data, 4. Results, and 5. Conclusions**.**

## Subsection heading

Subsection headings are left-justified, 10-pt. bold. Capitalize the first word, acronyms, and proper nouns.

## Paragraphs

Add a blank line above and below section headings and between paragraphs. Avoid headings or one-line paragraphs at the top or bottom of a page by using page breaks or extra blank lines. One method to preserve page breaks is to make the bottom margin a little larger than the specifications. Indentation is optional. This spacing is automatic if you are using the SPIE "section" styles in Word.

## Text

Text is 10 point and justified. Each sentence ends with a period and a single space before the next sentence.

## Figures and captions

Figures are centered. Use or insert .jpg, .tiff, or .gif illustrations instead of PowerPoint or graphic constructions. Captions go below figures. Indent 5 spaces from left margin and justify.



Figure . Figure captions are indented 5 spaces and justified. If you are familiar with Word styles, you can insert a field code called Seq figure which automatically numbers your figures.

## Tables and captions

Tables are centered. The caption goes above the table. The caption text should indent 5 spaces from left margin and justify. Table examples are on page 1 and 2 and below.

# Multimedia figures – video and audio files

Video and audio files can be included for publication. Table 3 lists the specifications for the mulitimedia files. Use a screenshot or another .jpg illustration for placement in the text. Use the file name to begin the caption. The text of the caption must end with the text "<http://dx.doi.org/doi.number.goes.here>" which tells the SPIE editor where to insert the hyperlink in the digital version of the manuscript.

Here is a sample illustration and caption for a multimedia file:



Video . A label of “Video/Audio 1, 2, …” should appear at the beginning of the caption to indicate to which multimedia file it is linked . Include this text at the end of the caption: <http://dx.doi.org/doi.number.goes.here>

Table 3. Information on video and audio files that can accompany a manuscript submission.

|  |  |  |
| --- | --- | --- |
| **Item** | **Video** | **Audio** |
| File name | Video1, video2, … | Audio1, audio2, … |
| Number of files | 0-10 | 0-10 |
| Size of each file | max. 5 mb | max. 5mb |
| File types accepted | .mpeg, .mov (Quicktime), .wmv (Windows Media Player) | .wav, .mp3 |

# Mathematical Equations

Use common fonts like Times Roman in your math equations. A math reference in a paragraph sentence such as  is not numbered. The steps of a mathematical argument can be numbered using a right-aligned tab for clarity, for example

α =  (1)

and

*ρ*=. (2)

# using THIS TEMPLATE AND ITS Automatic formatting

This document template and more information is available on SPIE.org:

<http://spie.org/x14101.xml> or by emailing [authorhelp@spie.org](mailto:authorhelp@spie.org)

## View the pre-formatted styles

To see the formats available with this manuscript, go to the Format menu and choose "Styles and Formatting". To view which style is being used in any part of this document, place your cursor on the line and look in the Styles and Formatting display.

## Using SPIE styles

To use this template in Microsoft Word: open this file and save it to a new file name. Type over the existing text with your paper. Alternately, you can delete all text, and select the SPIE style for each paragraph from the Styles and Formatting menu.

## Notes on SPIE styles

The styles listed in Table 2 automatically add extra spacing before and/or after paragraphs: SPIE title, SPIE authors-affiliations, SPIE section heading, SPIE subsection heading, and SPIE body text. The 1.1 Heading 2 style automatically goes into the body text style after one paragraph return.

reference linking

SPIE is able to display the references section of your paper in the SPIE Digital Library, complete with links to referenced journal articles, proceedings papers, and books, when available. This added feature will bring more readers to your paper and improve the usefulness of the SPIE Digital Library for all researchers.

Denote reference citations within the text of your paper by means of a superscript number. List references at the end of the paper in numerical order, and enclose the reference number in square brackets. Include the following information (as applicable).

If you use this formatting, your references will link your manuscript to other research papers that are in the CrossRef system. Exact punctuation is required for the automated linking to be successful.

book: [1] Booth, N. and Smith, A. S., [Infrared Detectors], Goodwin House Publishers, New York & Boston, 241-248 (1997).

**journal paper:** [2] Davis, A. R., Bush, C., Harvey, J. C. and Foley, M. F., "Fresnel lenses in rear projection displays," SID Int. Symp. Digest Tech. Papers 32(1), 934-937 (2001).

**proceedings paper:** [3] Van Derlofske, J. F., "Computer modeling of LED light pipe systems for uniform display illumination," Proc. SPIE 4445, 119-129 (2001).

website: [4] Myhrvold, N., “Confessions of a cybershaman,” Slate, 12 June 1997, <http://www.slate.com/CriticalMass/97-06-12/CriticalMass.asp> (19 October 1997).

-OR-

[4] Myhrvold, N., “Confessions of a cybershaman,” Slate, 12 June 1997, http://www.slate.com/CriticalMass/97-06-12/CriticalMass.asp (19 October 1997).

References

1. Booth, N. and Smith, A. S., [Infrared Detectors], Goodwin House Publishers, New York & Boston, 241-248 (1997).
2. Davis, A. R., Bush, C., Harvey, J. C. and Foley, M. F., "Fresnel lenses in rear projection displays," SID Int. Symp. Digest Tech. Papers 32(1), 934-937 (2001).
3. Van Derlofske, J. F., "Computer modeling of LED light pipe systems for uniform display illumination," Proc. SPIE 4445, 119-129 (2001).
4. Myhrvold, N., “Confessions of a cybershaman,” Slate, 12 June 1997, <http://www.slate.com/CriticalMass/97-06-12/CriticalMass.asp> (19 October 1997). [www.optics4yurresearch.com/7752.html](http://www.optics4yurresearch.com/7752.html)

[5] Jones, C. J., Director, Miscellaneous Optics Corporation, interview, Sept. 23 2011

[6] FamilyName, GivenName Initial., "Title," Source, pg# (year).