3D reconstruction from correspondences derived from perceived motion

Richard Kirby\*, Ross Whitaker

University of Utah, School of Computing, 50 S. Central Campus Drive Room 3190, SLC UT, USA, 84112

**Abstract**. 3D reconstruction from image pairs taken from different camera perspectives relies on finding corresponding points between the images and using those corresponding points to estimate a dense disparity map. Today's correspondence finding algorithms primarily use image features or pixel intensities common between image pairs. Some 3D computer vision applications, however, don't produce the desired results using correspondences derived from image features or pixel intensities. Two examples are the multimodal camera rig and the center region of a coaxial camera rig. In this paper we present an image correspondence finding technique that aligns image sequences using optical flow fields. Our method applies to applications where there is inherent motion between the camera rig and the scene and where the scene has enough visual texture to produce optical flow. We apply the technique to a traditional binocular stereo rig consisting of an RGB/IR camera pair and to a coaxial camera rig. We present results for synthetic flow fields and for real images sequences. With real image sequences we achieve flow field alignment of better than 1/10 of a pixel and reconstruction of relative camera/scene velocity of ??.

**Keywords**: image correspondence finding, 3D reconstruction, stereo endoscope, stereo borescope, multi-modal, graph cuts.

**\***Richard Kirby**,** E-mail: [richard.kirby@utah.edu](mailto:myemail@university.edu)

# 1 Introduction

Finding corresponding points in image pairs taken from two different perspectives is one of the most active areas of research in computer vision [16, 32, 34]. It forms the basis of 3D reconstruction as well as being a critical component of many other computer vision and image processing applications that require pixel by pixel alignment between image pairs [15, 41].

Existing correspondence finding techniques are based on matching pixel intensity values or features which are derived from pixel intensity values. This, in turn, allows the estimation of dense disparity maps which, given the camera geometry, allows the estimation of dense depth maps. Where image registration is the ultimate goal, the correspondences provide the geometrical transformations that allows one image, the sensed image, to be transformed into the second image, the reference image.

There are computer vision applications, however, where traditional correspondence finding techniques do not produce the desired results. Two notably cases are multimodal camera rigs where the images produced from different sensor types are not similar enough to be aligned using pixel intensities or features [41] and the center region of a coaxial camera rig [28] where the disparity is too small to produce good triangulation. There are also multi-camera applications where it is desirable to augment the use of pixel intensities and/or image features to improve the finding of intra-camera correspondences.

In this paper we introduce a novel automated method for finding correspondences using the optical flow fields from two cameras. We apply the technique to images acquired by a multimodal stereo rig where one camera contains an RGB sensor and the other camera contains an IR sensor, as well as to a coaxial camera rig. In applications where there is sufficient motion between the camera rig and the scene (scanning security camera, camera mounted on a vehicle, endoscope, etc.) and where the scene exhibits enough texture to produce optical flow, our method finds correspondences between multi-view image sequences without using intra-camera pixel intensities or features. From these correspondences we estimate dense disparity maps with accuracies similar to, and in certain cases, substantially better than, techniques that align images based on image features or pixel intensities.

This paper extends the work previously presented in two conference publications [19, 20] that solved the energy minimization problem using variational methods. In this paper we apply graph-cuts to solve the energy minimization problem, add additional support to the derivations, and show the results on a wider range of real images.

# 2 Related Work

## 2.1 Multimodal Binocular Stereo

Aligning images from stereo rigs consisting of cameras with multimodal sensors has been an active research area for the last decade and a half. Initially inspired by the work done to match medical images to models [38] it has more recently been motivated by the need for surveillance systems that use a combination of visible light and infrared cameras to detect targets. As noted by Yaman and Kalkan [39], traditional image alignment techniques used in stereo vision are not applicable to multimodal camera rigs because the pixel intensities can be substantially different in a visible light image vs. an IR image. Solutions to the multimodal problem fall into two broad categories. The first uses Mutual Information (MI). MI was original proposed by Viola and Wells [38] to match medical images to models. To our knowledge, Egnal [10] was the first to use MI as a similarity measure to match multimodal stereo images. Since then, numerous improvements have been made including adaptive windowing [11], incorporating prior probabilities [12], regions of interest [23-25], and extending MI using gradient information [8].

More recently, local self similarity (LSS), originally used in template matching, was proposed for use in a multimodal camera rig [35]. Most recently Yaman and Kalkan [39] used MI to generate dense disparity maps from multimodal camera rigs.

The method we present avoids using visual similarity measures between the images from the two different sensor types by computing the optical flow fields from the two sensors and then aligning the flow fields. This permits images with no common features to be aligned as long as there is motion between the camera and the scene and the scene has enough texture to produce optical flow.

Verri and Poggio [37] have shown that in many cases optical flow is not equivalent to the motion field. While optical flow algorithms have improved substantially since the Verri and Poggio paper (see [33] and [7] for summaries of the progression of optical flow algorithm development); optical flow errors caused by the aperture problem, non-Lambertian surfaces, and non-uniform changing illumination, still exist.

For finding image correspondences, however, the optical flow fields do not need to be equivalent to the motion fields. For example, errors caused by the aperture problem where only the motion tangential to edges is detected or errors caused by moving shadows, will be perceived by the two sensors identically and alignment is unaffected. The primary requirement is that the optical flow computation be invariant to different light wavelengths.

(a)

(b)

Fig. 1 Multimodal camera rig image pair: (a) IR and (b) RGB.

## 2.2 Depth from Zooming - Coaxial Camera

Depth from images taken at different focal lengths along a common optical axis was first proposed by Ma and Olsen [28]. Lavest et al. [27, 26] provide a proof for inferring 3D data from images taken at multiple focal lengths 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. [13] present a distance measurement system for mobile robots using zooming. Most recently, Zhang and Qi [40] describe a method for 3D reconstruction from multi-focal length images using a snake-search algorithm.

The primary reason researchers have focused on using a single camera at different focal lengths to do 3D reconstruction has 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 than a binocular stereo rig.

The coaxial camera rig [18] 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 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.

# 3 Energy Formulation

Make some comments here about relating the flow from one camera to the flow from the other camera and that we will show how this is done for two camera rig types.

## 3.1 Binocular Stereo

Referring to Figure 2, let , represent points in the image domain of the left and right cameras. Let the disparity between and such that and represent the same point in the scene. Let the focal lengths of the cameras and the distance between the optical center of the left camera and a point in the scene corresponding to at time and , the distance being measured along the optical axis. is then the difference along the Z axis for each point between and . the distance from the optical axis to a point in the scene and the change in the distance from the optical axis between time and . the stereo baseline. the projection of the 3D motion (the ideal flow) of point in the scene onto the image planes in the left and right cameras.

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.

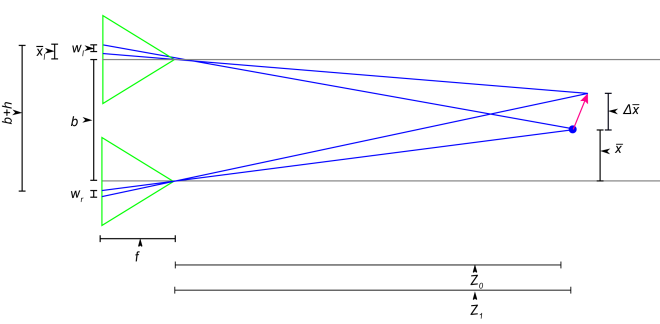


Fig. 2 Binocular stereo camera rig geometry.

Solving for the ratio of the two flows at corresponding points gives:

(3)

where:

(4)

has a physical interpretation. From equation (4) it can be seen that if . 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 along the image axis. compensates for this parallax and can be solved for directly from the camera geometry.

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

(5)

The second term is a smoothness term:

(6)

The total energy that we want to minimize is:

(7)

where and are tuning constants.

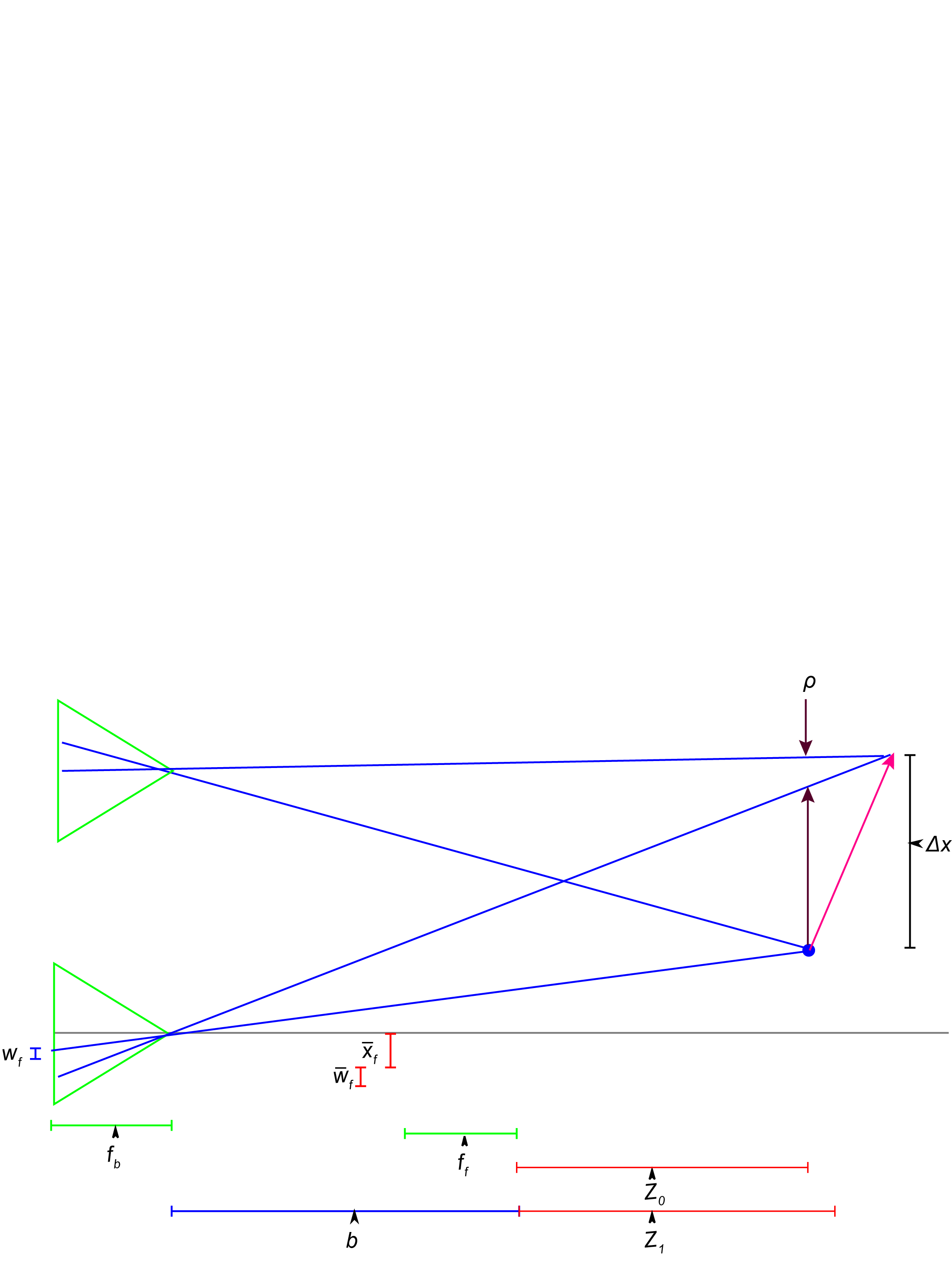


Fig. 3 Parallax caused by start and end points having a different depth.

## 3.2 Dual Focal Length Binocular Stereo Epipolar Geometry

## 3.3 Coaxial Camera Rig

Referring to Figure 1, 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 corresponding 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

Fig. 4 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 2 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:

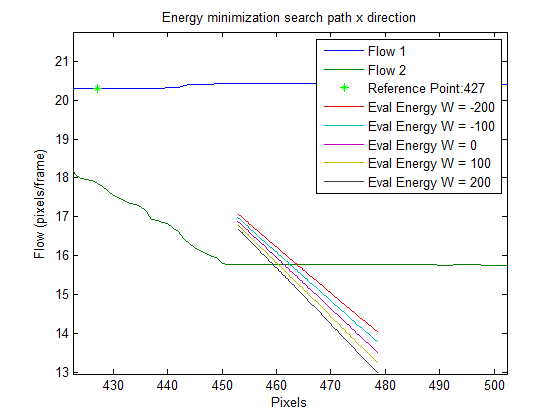
(8)

where and are tuning constants.

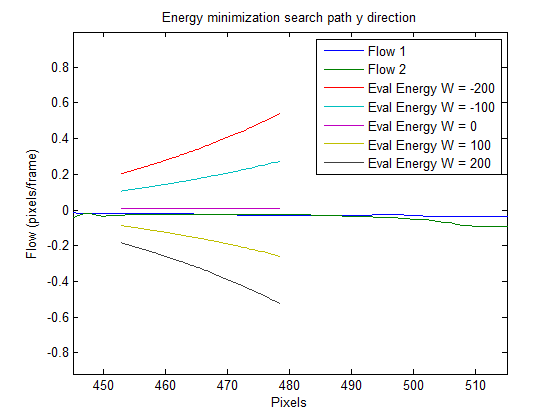
Observation to work in with some plots:

1. The matching term matches both the flow in the x and y direction, which reduces the chance of a false positive match.
2. We are solving for Z and W directly, which allows sub-pixel disparity.
3. There is a 3D matching. As Z and W vary, the energy in x direction and y direction vary and in theory only one point matches. Need to try this along a line.

Here is an example on a frontal planar surface, which for optical flow is analogous to a featureless region in correspondences from features. We have motion in the X direction, no motion in the Y or Z direction. Looking at the optical flow in x, there is a Z estimate which gives a minimum energy for each value of W.



However, looking at the flow in the y direction (which is zero) there is only one value of Z which produces an energy minimum.



# 4 Numerical Solution

Equations 7? and 8? define global energies, which can be solved using a variety of algorithms including variational methods [19, 20], simulated annealing [3, 14, 30], cooperative methods [29, 33] and more recently graph-cut algorithms [6, 4, 17, 31, 33]. In our previous work [19, 20] we used a variational methods approach, which was slow and had some difficulties with the stopping condition. In this paper we solve the energy minimization using the graph-cuts algorithm proposed by Boykov et. al [5, 4, 9, 21] and implemented by Veksler and Delong [36], which partitions the image pixels and labels into two sets along a surface (work on this partitioning idea).

Using Veksler and Delong's notation:

(8)

where

(8)

is an arbitrary finite set of labels

is the sum over all pixels in the reference image of the cost of assigning label to pixel . Where is a specific (Z, W) combination and is a specific pixel location. Our matching term penalizes the difference between the optical flow in the reference image at pixel and the optical flow in the sensed image at (. This is a two component penalty as both components of the optical flow contribute to the cost. This term is equivalent to our match term.

is the sum over all pairs of neighboring pixels in the reference image, where are 4-connected. This cost defines the pixel neighborhood structure and penalizes neighboring pixels that have different labels.

is the sum over all subsets of labels. This cost defines the label structure and penalizes non-adjacent Z and W values. This terms allows us to define the neighborhood structure of the labels where adjacent labels are not adjacent.

The combination of Esmooth and Elabel produce an 8-connected smoothness term similar to our Esmooth. 4 of the 8 connections are spatial, two encourage adjacent Z values and two encourage adjacent W values.

We use the expansion algorithm for energy minimization because 3 sets of labels

Our global energy has two main differences versus a traditional binocular stereo energy: 1) in matching optical flow, each pixel location in the reference frame has two values, one for the optical flow in the x direction and a second for the optical flow in the y direction and 2) instead of solving for disparity, we solve for W and Z directly and simultaneously.

We use the graph cuts algorithm developed by Boykov et. al. [4] that minimizes energies of the type of equations 7? and 8? for an arbitrary finite set of labels by partitioning the image pixels

As in [22], we define as the set of pixels in the left and right images. Put some kind of general definition for vertices and nodes/edges and how the energy relates to the edges. Make sure script V is defined. Also put some notes in about discrete increments for Z and W similar to stereo's discrete increments for disparity. is the set of pairs of pixels (one from the left image and one from the right image) that might correspond. Then:

(8)

Where is the set of all possible combinations of Z and W: . Both Z and W can be in any units of measurement, which allows subpixel flow alignment. Similar to stereo graph-cuts implementations that assume that disparity falls within some limited range, we assume that Z and W fall within a limited range.

# 5 Results

# 6 Conclusions

## 1.1 Use of This Template

Authors may use this Microsoft (MS) Word template by employing the relevant styles from the Styles and Formatting list (which is accessed from the Styles group in the Home ribbon, Fig. 1):

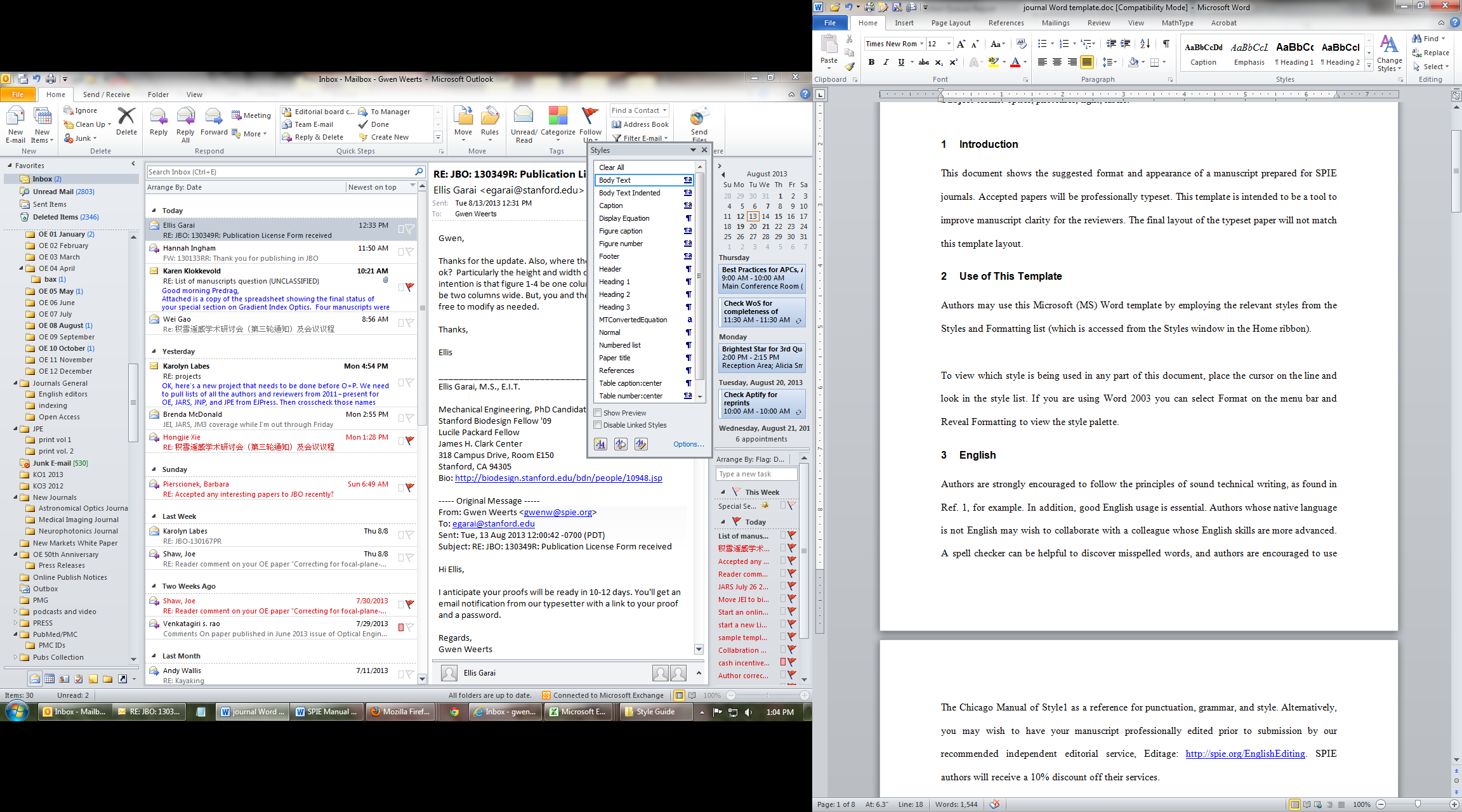


Fig. 1 MS Word 2010 Home ribbon. Red arrow indicates where to access Styles window.

To view which style is being used in any part of this document, place the cursor on the line and look in the style list (Fig. 2):

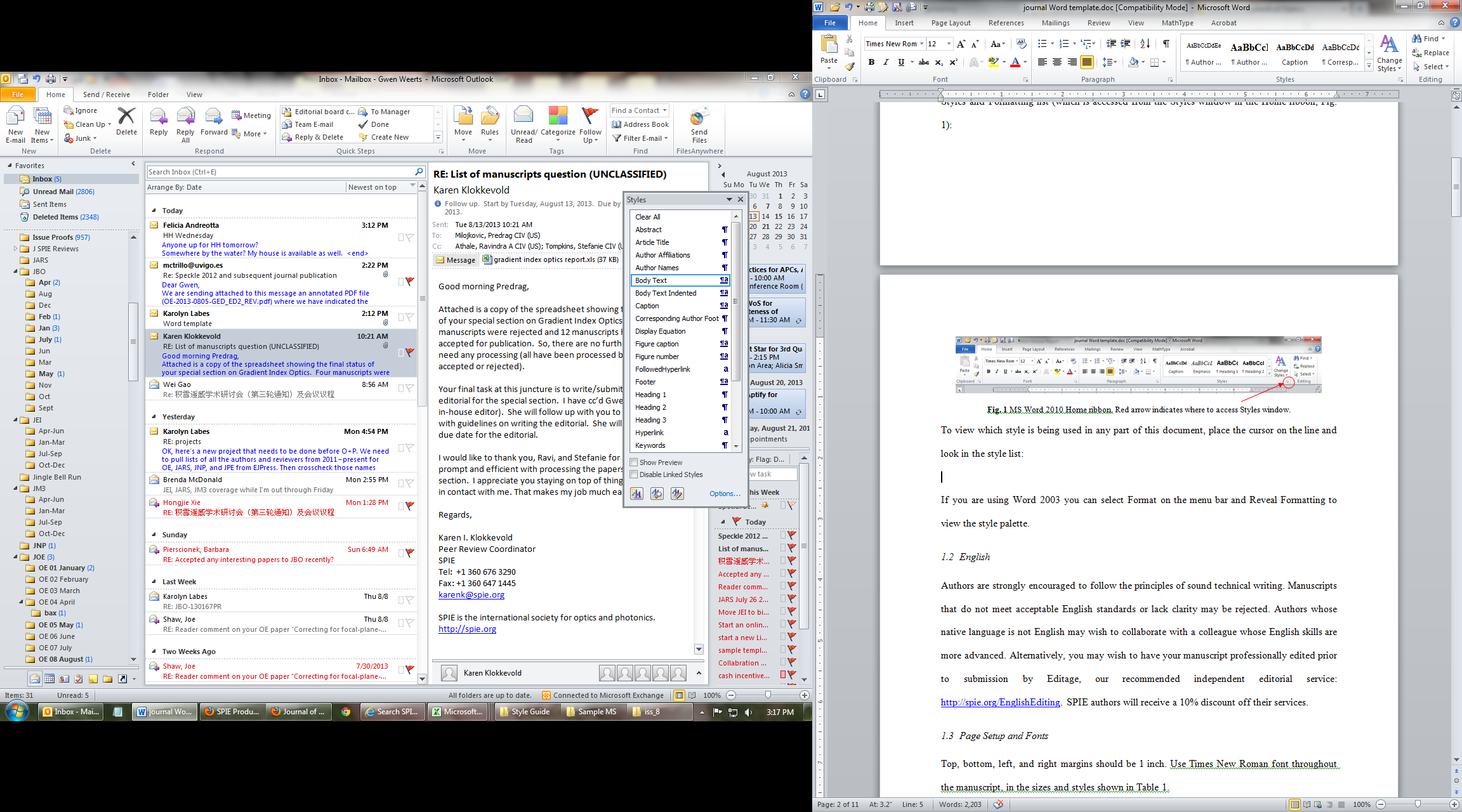


Fig. 2 Styles window.

If you are using Word 2003 you can select Format on the menu bar and Reveal Formatting to view the style palette.

## 1.2 English

Authors are strongly encouraged to follow the principles of sound technical writing. Manuscripts that do not meet acceptable English standards or lack clarity may be rejected. Authors whose native language is not English may wish to collaborate with a colleague whose English skills are more advanced. Alternatively, you may wish to have your manuscript professionally edited prior to submission by Editage, our recommended independent editorial service: <http://spie.org/EnglishEditing>. SPIE authors will receive a 10% discount off their services.

## 1.3 Page Setup and Fonts

Top, bottom, left, and right margins should be 1 inch. Use Times New Roman font throughout the manuscript, in the sizes and styles shown in Table 1.

**Table 1** Recommended fonts and sizes.

|  |  |
| --- | --- |
| **Style name** | **Brief description** |
| **Article Title** | **16 pt, bold** |
| **Author Names** | **12 pt, bold** |
| Author Affiliations | 10 pt |
| Abstract | 10 pt |
| Keywords | 10 pt |
| **Heading 1** | **12 Pt, bold** |
| *Heading 2* | *12 pt, italic* |
| *Heading 3* | *11 pt, italic* |
| Body Text | 12 pt |
| Figure caption | 10 pt |
| Table caption | 10 pt |

# 2 Parts of a Manuscript

This section describes the normal structure of a manuscript and how each part should be handled.

## 2.1 Title and Author Names

The article title appears left justified at the top of the first page. The title font is 16 pt, bold. The rules for capitalizing the title are the same as for sentences; only the first word, proper nouns, and acronyms should be capitalized. Do not begin titles with articles (e.g., a, an, the) or prepositions (e.g., on, by, etc.). The word “novel” should not appear in the title, as publication will imply novelty. Avoid the use of acronyms in the title, unless they are widely understood. Appendix A contains more about acronyms.

The list of authors immediately follows the title. The font is 12 pt bold and the author names are left justified. The author affiliations and addresses follow the names, in 10 pt normal font and left justified. For multiple affiliations, each affiliation should appear on a separate line. Superscript letters (a, b, c, etc.) should be used to associate multiple authors with their respective affiliations. The corresponding author should be identified with an asterisk, and that person's email address should be provided below the keywords.

## 2.2 Abstract

The abstract should be a summary of the paper and not an introduction. Because the abstract may be used in abstracting journals, it should be self-contained (i.e., no numerical references) and substantive in nature, presenting concisely the objectives, methodology used, results obtained, and their significance. It should be 200 words, maximum. For further guidelines, please read the brief article titled "[How to Write an Abstract (PDF)](http://spie.org/Documents/Publications/How%20to%20Write%20an%20Abstract.pdf)," by Philip Koopman. (Courtesy of Philip Koopman, Carnegie Mellon University.)

## 2.3 Subject Terms/Keywords

Keywords are required. Please provide 3-6 keywords related to your paper.

## 2.4 Body of Paper

The body of the paper consists of numbered sections that present the main findings. These sections should be organized to best present the material.

It is often important to refer back (or forward) to specific sections. Such references are made by indicating the section number, for example, “In Sec. 2 we showed…” or “Section 2.1 contained a description….” If the word Section, Reference, Equation, or Figure starts a sentence, it is spelled out. When occurring in the middle of a sentence, these words are abbreviated Sec., Ref., Eq., and Fig.

At the first occurrence of an acronym, spell it out followed by the acronym in parentheses, e.g., charge-coupled diode (CCD).

## 2.5 Footnotes

Use textual footnotes only when necessary to present important documentary or explanatory material whose inclusion in the text would be distracting.[[1]](#footnote-1)

## 2.6 Appendices

SPIE journals do not accept supplementary materials. However, it is acceptable to include an Appendix when necessary, such as derivations of equations, proofs of theorems, and details of algorithms. Equations and figures appearing in Appendices should continue sequential numbering from earlier in the paper.

## 2.7 Disclosures

Conflicts of interest should be declared under a separate header, above Acknowledgments. If the authors have no competing interests to declare, then a statement should be included declaring no conflicts of interest. For assistance generating a disclosure statement, see the form available from  the International Committee of Medical Journal Editors website: <http://www.icmje.org/conflicts-of-interest/>

## 2.8 Acknowledgments

Acknowledgments and funding information should be added after the conclusion, and before references. Include grant numbers and the full name of the funding body. The acknowledgments section does not have a section number.

## 2.8 References

The References section lists books, articles, and reports that are cited in the paper. The references are numbered in the order they are cited. Examples of reference styles are given at the end of this template.

For books1,2 the listing includes the list of authors (initials plus last name), book title (in italics), page or chapter numbers, publisher, city, and year of publication. Journal article references3,4 include the author list, title of the article (in quotes), journal name (in italics, properly abbreviated), volume number (in bold), inclusive page numbers or citation identifier, and year. A reference to a proceedings paper or a chapter in an edited book5,6 includes the author list, title of the article (in quotes), conference name (in italics), editors (if appropriate), volume title (in italics), volume number if applicable (in bold), inclusive page numbers, publisher, city, and year.

The references are numbered in the order of their first citation. Citations to the references are made using superscripts, as demonstrated in the preceding paragraph. One may also directly refer to a reference within the text, for example, “as shown in Ref. 5 ...” Two or more references should be separated by a comma with no space between them. Multiple sequential references should be displayed with a dash between the first and last numbers.1–5

### 2.8.1 Reference linking and DOIs

A Digital Object Identifier (DOI) is a unique alphanumeric string assigned to a digital object, such as a journal article or a book chapter, that provides a persistent link to its location on the internet. The use of DOIs allows readers to easily access cited articles. Authors should include the DOI at the end of each reference in brackets, if a DOI is available. See examples at the end of this manuscript. A free DOI lookup service is available from CrossRef at <http://www.crossref.org/freeTextQuery/>. The inclusion of DOIs will facilitate reference linking and is highly recommended.

## 2.9 Author Biographies and Photographs

Author biographies are requested but not required. Biographies are not to exceed 75 words. Biographies should be placed at the end of the manuscript. Personal information such as hobbies or birthplace/birthdate should not be included.

# 3 Section Formatting (Heading 1)

Type each section heading on a separate line using the appropriate style from the style list. Sections should be numbered sequentially.

Paragraphs that immediately follow a section heading are leading paragraphs and should not be indented, according to standard publishing style. The same goes for leading paragraphs of subsections and sub-subsections. In this MS Word template, use the Body Text style for leading paragraphs and the Body Text Indented style for all subsequent paragraphs.

## 3.1 Subsection Headings (Heading 2)

Subsection headings should be numbered 1.1, 1.2, etc.

### 3.1.1 Sub-subsection headings (Heading 3)

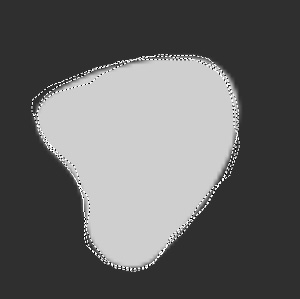
Sub-subsection headings should be numbered 1.1.1, 1.1.2, etc. Only the first word is capitalized.

# 4 Figures and Tables

## 4.1 Figures

Figures are numbered in the order in which they are called out in the text. It is acceptable to embed figures in the manuscript, but all figures must also be submitted as individual files in .tif, .eps, .png, PostScript, or PDF format. We cannot accept application files, i.e., Corel Draw, Microsoft Word, PowerPoint, etc. All figure parts must be labeled (a), (b), etc. Each figure file should contain all parts of the figure. For example, Fig. 3 contains two parts (a) and (b); therefore, all of the labeled parts should be combined in a single file for Fig. 3. Further details about figure formatting can be found at <http://spie.org/x85020.xml#Artwork>.

A figure caption list should be provided after the references.

sy00590_

(b)

(a)

Fig. 3 Example of a figure caption: (a) sun and (b) blob.

## 4.2 Tables

Tables are numbered in the order in which they referred to. They should appear in the document in numerical order and as close as possible to their first reference in the text. Table captions are handled identically to those for figures, except that they appear above the table, and when called out in the text, the word “Table” is always spelled out. See Table 1 for an example.

## 4.3 Multimedia

Please refer to the multimedia guidelines at <http://spie.org/x85020.xml#Multimedia> for specific submission guidelines and requirements. The following types of multimedia files are accepted: QuickTime Non-Streaming video (.qt or .mov), MPEG (.mpg or .mp4). The recommended maximum size for each multimedia file is 10-12 MB. Authors must insert a representative “still” image from the video file in the manuscript as a “figure.” This still image will be linked by the publisher to the actual video file, as will the caption label. Video files should be named Video 1, Video 2, etc. Otherwise multimedia files are treated in the same manner as figures. The multimedia file type should be included in parentheses at the end of the figure caption, along with the file size. See Video 1 for an example.



Video 1 Example of a multimedia still image (MPEG, 2.5 MB).

# Appendix A: Miscellaneous Formatting Details

## A.1 Formatting Equations

Equations may appear inline with the text, if they are simple, short, and not of major importance; e.g., a = b/c. Important equations appear on their own line. For example, “The expression for the field of view is

, (1)

where *a* is the …” Principal equations are numbered, with the equation number placed within parentheses and right justified. Authors are strongly encouraged to use MS Word Equation Editor or MathType to create both in-text and display equations. Equations are considered to be part of a sentence and should be punctuated accordingly.

## A.2 Formatting Theorems

To include theorems in a formal way, the theorem identification should appear in a 10-pt, bold font, left justified and followed by a period. Formal statements of lemmas and algorithms receive a similar treatment. The text of the theorem continues on the same line in normal, 10-pt font. For example,

**Theorem 1.** For any unbiased estimator…

## Disclosures

Conflicts of interest should be declared under a separate header. If the authors have no relevant financial interests in the manuscript and no other potential conflicts of interest to disclose, a statement to this effect should also be included in the manuscript.

## Acknowledgments

This unnumbered section is used to identify people who have aided the authors in accomplishing the work presented and to acknowledge sources of funding.

## References

1. G. J. Alred, C. T. Brusaw, and W. E. Oliu, *Handbook of Technical Writing*, 7th ed., St. Martin’s, New York (2003).
2. L. C. Perelman, J. Paradis, and E. Barrett, *Mayfield Handbook of Technical and Scientific Writing*, Mayfield, Mountain View, California (1997).
3. A. Harris et al., “Free-space optical wavelength diversity scheme for fog mitigration in a ground-to-unmanned-aerial-vehicle communications link,” *Opt. Eng.* **45**(8), 086001 (2006) [[doi:](http://spiedl.aip.org/jhtml/doi.jsp)10.1117/1.2338565].
4. N. Metropolis et al., “Equations of state calculations by fast computing machine,” *J. Chem. Phys****.* 21**(6), 1087-1091 (1953).
5. S. F. Gull, “Developments in maximum-entropy data analysis,” in *Maximum Entropy and Bayesian Methods*, J. Skilling, Ed., pp. 53-71, Kluwer Academic, Dordrecht, Netherlands (1989).
6. K. M. Hanson, “Introduction to Bayesian image analysis,” in Image Processing, M. H. Loew, Ed., *Proc. SPIE* **1898**, 716-731 (1993) [[doi:](http://spiedl.aip.org/jhtml/doi.jsp)10.1117/12.154577].

**First Author** is an assistant professor at the University of Optical Engineering. He received his BS and MS degrees in physics from the University of Optics in 1985 and 1987, respectively, and his PhD degree in optics from the Institute of Technology in 1991. He is the author of more than 50 journal papers and has written three book chapters. His current research interests include optical interconnects, holography, and optoelectronic systems. He is a member of SPIE.  
  
Biographies and photographs for the other authors are not available.

**Caption List**

Fig. 1 MS Word 2010 Home ribbon. Red arrow indicates where to access Styles window.

**Fig. 2** Styles window.

**Fig. 3** Example of a figure caption: (a) sun and (b) blob.

Video 1 Example of a multimedia still image (MPEG, 2.5 MB).

**Table 1** Recommended font sizes and styles.

[1] N. Asada, m. Baba, and A. Oda, "Depth from Blur by Zooming," in *Proceedings of the Vision Interface Annual Conference*, Ottawa, Canada, 2001.

[2] M. Baba, N. Asada, and T. Migita, "A Thin Lens Based Camera Model for Depth Estimation from Defocus and Translation by zooming," in *Proc. 15th International Conference on Vision Interface*, Calgary, Canada, 2002.

[3] S. T. Barnard, "Stochastic stereo matching over scale," *IJCV,* vol. 3(1):17-32, 1989.

[4] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998.

[5] Y. Boykov and V. Kolmogorov, "An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision," in *IEEE TPAMI*, 2004.

[6] Y. Boykov and V. Kolmogorov, "Graph Cuts in Vision and graphics theories and Applications," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. Vol. 26, No. 9, September, 2004.

[7] T. Brox and J. Malik, "Large Displacement Optical Flow Desriptor Matching in Variational Motion Estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 2010.

[8] F. B. Campo, R. L. Ruiz, and A. D. Sappa, "Multimodal Stereo Vision System: 3D Data Extraction and Algorithm Evaluation," *IEEE Journal of Selected Topics In Signal Processing, vol. 6, no.5,* 2012.

[9] A. Delong and H. Osokin, "Fast Approximate Energy Minimization with Label Costs," in *CVPR*, 2010.

[10] G. Egnal, "Mutual information as a Stereo Correspondence Measure," *Technical Report MS-CIS-00-20, Computer and Information Science, University of Pennsylvania, Philadelphia, USA,* 2000.

[11] C. Fookes, A. Lamanna, and M. Bennamoun, "A new stereo image matching technique using mutual information," in *In Proceedings of the International Conference on Computer, Graphics and Imaging, CGIM'01, pages 168-173, Honolulu, USA, 2001. Iasted, ISBN 0-88986-303-2.*, 2001.

[12] C. Fookes, S. Maeder, S. Sridharan, and J. Cook, "Multi-Spectral Stereo Image Matching using Mutual Information," in *Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization, and Transmission*, 2004.

[13] H. Gao, J. Liu, Y. Yu, and Y. Li, "Distance measurement of zooming image for a mobile robot," *International Journal of Control, Automation and Systems,* vol. 11, pp. 782-789, 2013.

[14] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images.," *IEEE TPAMI,* vol. 6(6):721-741, 1984.

[15] A. A. Goshtasby, *Image registration principles tools methods*: Springer, 2012.

[16] R. Hartly and A. Zisserman, *Multiple View Geometry in computer vision*: Cambridge University Press, 2003.

[17] P.-J. Kack, "Robust Stereo Correspondence Using Graph cuts," PhD, KTH Numerical Analysis and Computer Sciences, 2004.

[18] R. Kirby, "Three Dimensional Surface Mapping System Using Optical Flow US2013321790A1," USA Patent, 2012.

[19] R. kirby and R. Whitaker, "3D reconstruction from images taken with a coaxial camera rig " in *SPIE Optics + Photonics*, San Diego, 2016.

[20] R. Kirby and R. Whitaker, "A novel automated method for doing revistration and 3D reconstruction from multi-modal RGB/IR image sequences," in *SPIE Optics + Photonics*, San Diego, 2016.

[21] V. Kolmogorov and R. Zabih, "What Energy Functions can be Minimized via Graph Cuts?," in *IEEE TPAMI*, 2004.

[22] V. Kolmogorov and R. Zabih, *Graph cut algorithms for binocular stereo with occlusions*: Springer, 2006.

[23] S. Krotosky and T. Mohan, "Registration of Multimodal Stereo Images using Disparity Voting from Correspondence Windows," in *IEEE Conf. on Advanced Video and Signal based Surveillance (AVSS’06)*, 2006.

[24] S. Krotosky and M. Trivedi, "Multimodal Stereo Image Registration for Pedestrian Detection," in *in Proc. IEEE Intell. Transp. Syst. Conf., Sep. 2006, pp. 109–114*, 2006.

[25] S. Krotosky and M. Trivedi, "Mutual information based registration of multimodal stereo videos for person tracking," *Computer Vision and Image Understanding,* vol. 106, pp. 270-287, 2007.

[26] J. Lavest, G. Rives, and M. Dhome, "Three Dimensional Reconstruction by Zooming," *IEEE Transactions on Robotics and Automation,* vol. 9, pp. 196-207, 1993.

[27] J. Lavest, G. Reves, and M. Dhome, "Modeling an Object of Revolution by Zooming," *IEEE Transactions on Robotics and Automation,* vol. VOL. II, NO. 2, April 1995, 1995.

[28] J. Ma and S. I. Olsen, "Depth from Zooming," *J. Opt. Soc. Am. A* vol. 7, October 1990 1990.

[29] D. Marr and T. Poggio, "Cooperative computation of stereo disparity," *Science* vol. 194, 283-287, 1976.

[30] J. Marroquin, S. Mitter, and T. Poggio, "Probabilistic solution of ill-posed problems in computational vision," *J. Am. Stat. Assoc.,* vol. 82(397):76-89, 1987.

[31] S. Roy and I. Cox, "A maximum-flow formulation of the N-camera stereo correspondence problem," *ICCV,* vol. pp. 492-499, 1998.

[32] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two frame stereo correspondance agorithms," in *IJCV*, 2001.

[33] D. Scharstein, R. Szeliski, and R. Zabih, "A taxonomy and evaluation of dense two frame stereo correspondence algorithms," *International Journal of Computer Vision* vol. 47, pp. 7-42, 2002.

[34] R. Szeliski, *Computer Vision. Algorithms and Applications*. New York: Springer, 2011.

[35] A. Toraby and G. Bilodeau, "Local self-similarity as a dense stereo correspondence measure for themal visible video registration," *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on,* pp. 61 - 67, 2011.

[36] O. Veksler and A. Delong. (2007). *GCoptimization*. Available: http://www.csd.uwo.ca/faculty/olga/software.html

[37] A. Verri and T. Poggio, "Motion Field and Optical Flow: Qualitative Properties," *IEEE Trans. Pattern Analysis and Machine Intelligence, 11(5), pp. 490-498, May,* 1989.

[38] P. Viola and W. M. Wells, "Alignment by Maximization of Mutual Information," *Intl. J. of Computer Vision, vol. 24, no. 2, pp. 137–154,,* 1997.

[39] M. Yaman and S. Kalkan, "An iterative adaptive multi-modal stereo-vision method using mutual information," *Journal of Visual Communication and Image Representation,* vol. 26, pp. 115-131, 2015.

[40] Y. Zhang and K. Qi, "Snake-Search Algorithm for Stereo Vision Reconstruction via Monocular System," presented at the The 5th Annual IEEE Conference on Cyber Technology in Automation, and Control, Intelligent Systems, Shenyang, China, 2015.

[41] B. Zitová and J. Flusser, "Image registration methods: a survey," *Image and Vision Computing,* vol. 21, pp. 977-1000, 2003.

1. Due to problems with HTML display, use of footnotes should be avoided. If absolutely necessary, the footnote mark must come at the end of a sentence. To insert a footnote, use the Insert menu, select Reference, then Footnote, change the number format to the style of asterisk, dagger, double-dagger, etc., and click OK. [↑](#footnote-ref-1)