

On Cross-Spectral Stereo Matching using Dense Gradient Features

Peter Pinggera^{1,2,3}
pinggera@alumni.tugraz.at

Toby Breckon¹
toby.breckon@cranfield.ac.uk

Horst Bischof²
bischof@icg.tugraz.at

¹ School of Engineering, Cranfield University, Bedfordshire, UK.

² Institute for Computer Graphics and Vision, TU Graz, Austria.

Here we address the problem of scene depth recovery within cross-spectral stereo imagery (each image sensed over a differing spectral range). We compare several robust matching techniques which are able to capture local similarities between the structure of cross-spectral images and a range of stereo optimisation techniques for the computation of valid dense depth estimates for this case.

As the performance of standard optical camera systems can be severely affected by environmental conditions the use of combined sensing systems operating in differing parts of the electromagnetic spectrum is increasingly common [5]. As a result, an attractive solution is the combination of both optical and thermal images in many sensing and surveillance scenarios as the complementary nature of both modalities can be exploited and the individual drawbacks largely compensated. Despite the inherent stereo setup of this common two sensor deployment, in practical scenarios it is rarely exploited. Here, we specifically deal with the recovery of dense depth information from thermal (far infrared spectrum) and optical (visible spectrum) image pairs where large differences in the characteristics of image pairs make this task significantly more challenging than the common stereo case (Figure 1A).

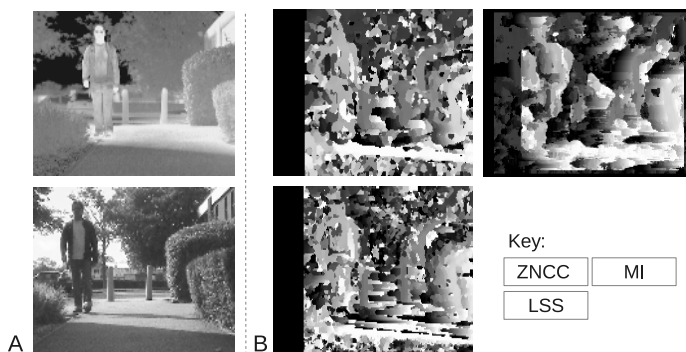


Figure 1: Performance of ZNCC, MI and LSS approaches

Prior work on cross-spectral stereo is weak and either recovers depth from isolated scene objects (Local Self-Similarity (LSS) features, [9]) or relies on an evaluation based on simulated cross-spectral imagery (Mutual Information (MI), [2, 3, 4]). Related work on the problem of radiometric differences in stereo image pairs [6] uses (amongst others) Zero Mean Normalised Cross Correlation (ZNCC). The poor performance of these prior techniques on an example cross-spectral stereo pair (Figure 1A) is shown in Figure 1B.

By contrast, we show cross-spectral stereo matching can be achieved, by using dense gradient features combined with strong optimisation criteria, to produce a scene depth image usable for further scene analysis and understanding (Figure 2). Our approach facilitates full scene depth recovery comparable in quality to standard optical stereo techniques under identical scene conditions.

This extends prior work which is limited to simulated cross-spectral results [2, 3, 4], or isolated object depth recovery [7, 8, 9]. We illustrate that dense gradient feature approaches outperform methods based on prior work using Mutual Information (MI) [2, 3, 4] and Local Self-Similarity (LSS) features [9]. Furthermore, we show that prior results on radiometric image differences [6] or simulated imagery [2, 3, 4] do not readily transfer to the true cross-spectral case.

The prevalence of dense gradient approaches, notably dense unsigned Histograms of Oriented Gradient (HOG) features [1], is shown over a range of disparity optimisation approaches with improved results under strong optimisation criteria of Graph Cuts (GC) and Semi-Global Matching (SGM). Although the results remain somewhat coarse in comparison to contemporary work in optical stereo [6], this work illustrates both :-

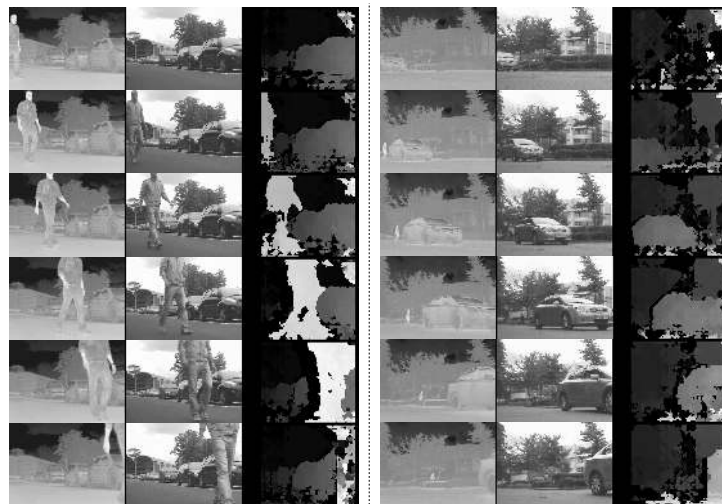


Figure 2: Two cross-spectral stereo sequences obtained using HOG+SGM depth recovery without explicit temporal consistency constraints

a) the additional challenge of cross-spectral stereo in comparison to other stereo matching cases (e.g. radiometric differences [6]) and b) that results suitable for further scene analysis and understanding are achievable via a dense gradient feature approach (Figure 2).

- [1] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 886–893. IEEE, 2005.
- [2] G. Egnal. Mutual information as a stereo correspondence measure. Technical report, University of Pennsylvania, 2000.
- [3] C. Fookes and S. Sridharan. Investigation & comparison of robust stereo image matching using mutual information & hierarchical prior probabilities. In *Proc. Second Int. Conf. on Signal Processing and Communication Systems*, pages 1–10. IEEE, 2008.
- [4] C. Fookes, A. Maeder, S. Sridharan, and J. Cook. Multi-spectral stereo image matching using mutual information. In *Proc. Second Int. Symposium on 3D Data Processing, Visualization and Transmission*, pages 961–968. IEEE, 2004.
- [5] A. Gaszczak, T.P. Breckon, and J.W. Han. Real-time people and vehicle detection from UAV imagery. In *Proc. SPIE Conf. Intelligent Robots and Computer Vision XXVIII: Algorithms and Techniques*, volume 7878, 2011.
- [6] H. Hirschmüller and D. Scharstein. Evaluation of stereo matching costs on images with radiometric differences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(9):1582–1599, 2009.
- [7] S. Krotosky and M. Trivedi. Mutual information based registration of multimodal stereo videos for person tracking. *Computer Vision and Image Understanding*, 106(2-3):270–287, 2007.
- [8] S. Krotosky and M. Trivedi. Registering multimodal imagery with occluding objects using mutual information: application to stereo tracking of humans. In R.I. Hammoud, editor, *Augmented Vision Perception in Infrared: Algorithms and Applied Systems*, chapter 14, pages 321–347. Springer, 2009.
- [9] A. Torabi and G.-A. Bilodeau. Local self-similarity as a dense stereo correspondence measure for thermal-visible video registration. In *IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 61–67, 2011.

³ Peter Pinggera is now with Daimler Research and Development, Germany.