

# Improved Depth Recovery In Consumer Depth Cameras via Disparity Space Fusion within Cross-spectral Stereo

Gregoire Payen de La Garanderie  
gregoire@hochet.info

Toby P. Breckon  
toby.breckon@durham.ac.uk

School of Engineering,  
Cranfield University, Bedfordshire, UK  
School of Engineering and Computing Sciences,  
Durham University, Durham, UK

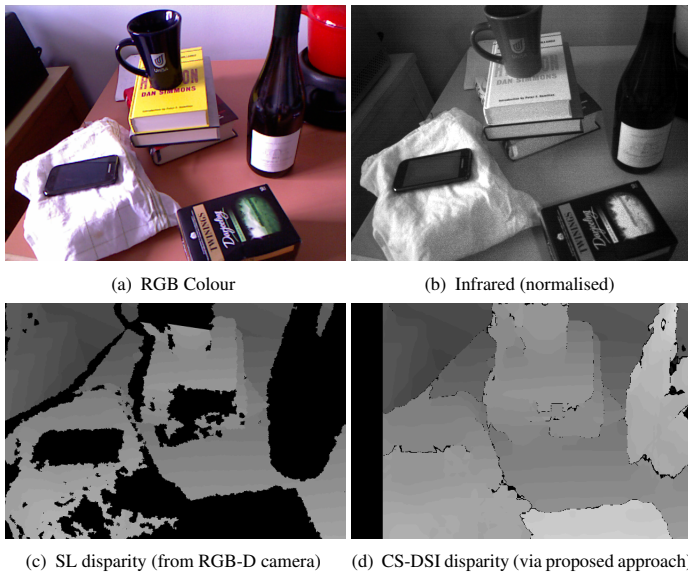


Figure 1: Fused disparity estimation:-  $(SL, CS) \rightarrow CS - DSI$

Low-cost consumer depth cameras have seen the combined use of colour and 3D depth most commonly leverage the use of near infrared structured light projection (830nm) with regular visible-band colour sensing (400-700nm) to provide co-registered colour (RGB) and depth (D) as combined RGB-D image components. The common physical characteristics of such devices - comprising a colour camera (Fig. 1(a)), an infrared pattern projector and corresponding infrared camera (Fig. 1(b)) (e.g. Microsoft Kinect / PrimeSense Carmine) - pose an obvious, yet commonly under-utilized, cross-modal stereo configuration.

Depth coverage in consumer depth cameras can be considerably improved based on the combined use of such cross-spectral stereo (CS) and near infrared structured light sensing (SL). Our joint approach, leveraging disparity information from both structured light and cross-spectral stereo, facilitates the recovery of global scene depth comprising both texture-less object depth, where stereo sensing commonly fails, and highly reflective object depth, where structured light active sensing commonly fails. The proposed solution is illustrated using dense gradient feature matching and is shown to outperform prior approaches [1, 2] that use late-stage fused cross-spectral stereo depth as a facet of improved sensing for consumer depth cameras.

We propose the use of "best in class" dense gradient features from [3] to facilitate recovery of secondary cross-spectral stereo disparity directly from the depth camera  $\{I_{ir}, I_{RGB}\}$  image pair. Our main contribution is the fusion of this secondary cross-spectral (CS) disparity information with *a priori* disparity information, obtained via conventional structured light (SL) sensing, within the disparity space image (DSI) constructed prior to conventional disparity optimization for scene depth recovery.

This is achieved by modifying the *disparity space image*, formed by  $C(x, y, d)$ , which constitutes the disparity cost space over which disparity optimization will be performed. We construct an alternative cost function,  $C_{DSI}(x, y, d)$  such that the use of disparity from structured light sensing,  $D_{SL}(x, y)$ , is incorporated as follows:

$$C_{DSI}(x, y, d) = \begin{cases} C_{HOG}(x, y, d) & \text{if } D_{SL}(x, y) \text{ is unavailable at pixel } (x, y) \\ low_c & \text{if } d = D_{SL}(x, y) \\ high_c & \text{if } d \neq D_{SL}(x, y) \end{cases} \quad (1)$$

Figure 1 shows both the disparity recovered by conventional structured light (SL) within such a device (Fig. 1(c)) and that recovered by our proposed cross-spectral disparity space image (CS-DSI) approach (Fig.

1(d)) for a given  $\{I_{ir}, I_{RGB}\}$  image pair (Fig. 1(a) / 1(b)).

Figure 2 shows the disparity results obtained from the  $\{I_{RGB}, I_{ir}, I_{depth}\}$  triplet shown in Fig. 2(a) - 2(c) for both the prior work of [1] (CS-union, Fig. 2(e)) and the proposed CS-DSI approach (Fig. 2(f)) against illustrative ground truth depth derived using manual depth labelling (Fig. 2(d)). The resulting CS-DSI disparity (Fig. 2(f)) presents a clearer disparity image with notably less missing disparity values and noise than CS-union (Fig. 2(e)) and the original SL disparity (Fig. 2(c)). This quality improvement resulting from CS-DSI is also present within Fig. 1.

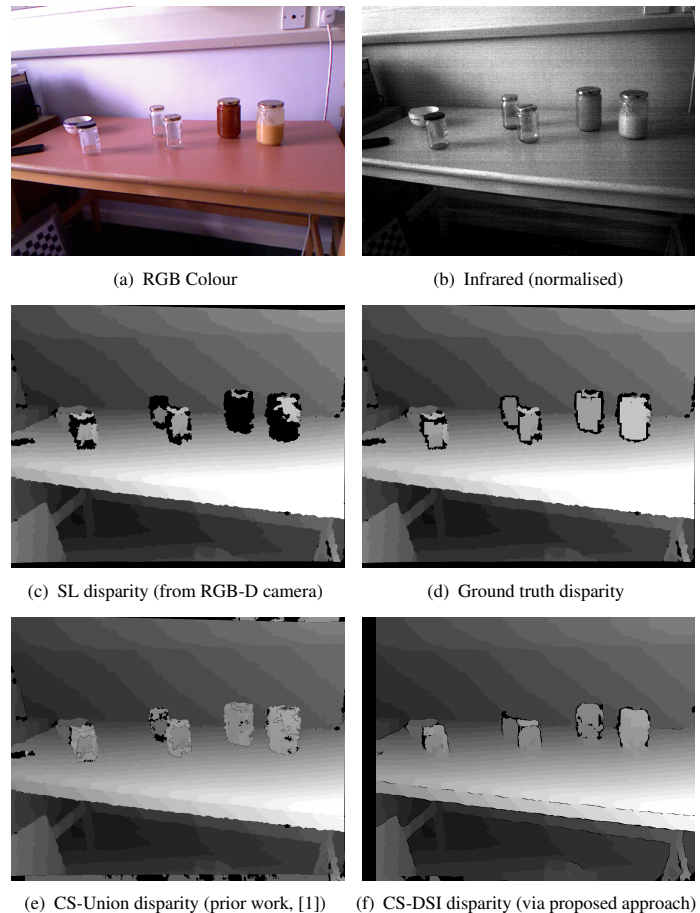


Figure 2: Disparity recovery on transparent and specular objects

Improved disparity can be recovered from a consumer depth camera based on the fusion of cross-spectral stereo and existing structured light sensing performed prior to conventional disparity space optimization. Missing depth information is recovered for transparent and specular objects in addition to that missing due to inter-object occlusions. This directly extends prior work [1, 2] which is shown to produce lesser depth recovery and requires computationally expensive scene dependant optimization.

- [1] W. C. Chiu, U. Blanke, and M. Fritz. Improving the Kinect by Cross-Modal Stereo. In *Proceedings of the British Machine Vision Conference*, pages 1–10, 2011. doi: 10.5244/C.25.116.
- [2] W. C. Chiu, U. Blanke, and M. Fritz. I spy with my little eye: Learning optimal filters for cross-modal stereo under projected patterns. *Proc. IEEE International Conference on Computer Vision Workshops*, pages 1209–1214, 2011. doi: 10.1109/ICCVW.2011.6130388.
- [3] P. Pinggera, T.P. Breckon, and H. Bischof. On cross-spectral stereo matching using dense gradient features. In *Proc. British Machine Vision Conference*, pages 526.1–526.12, September 2012. doi: 10.5244/C.26.103.