

MULTISPECTRAL INTEREST POINTS FOR RGB-NIR IMAGE REGISTRATION

D. Firmenich, M. Brown, S. Süsstrunk

School of Computer and Communication Sciences,
École Polytechnique Fédérale de Lausanne
email: damien.firmenich@epfl.ch

ABSTRACT

This paper explores the use of joint colour and near-infrared (NIR) information for feature based matching and image registration. In particular, we investigate multispectral generalisations of two popular interest point detectors (Harris and difference of Gaussians), and show that these give a marked improvement in performance when the extra NIR channel is available. We also look at the problem of multimodal RGB to NIR registration, and propose a variant of the SIFT descriptor that gives improved performance.

Index Terms— Interest points, Image Registration, Near Infrared

1. INTRODUCTION

Silicon is naturally sensitive in the near-infrared (750-1100nm) range, a fact that has been used in a number of compelling multimodal sensing approaches [1, 2, 3]. Applications of active near-infrared (NIR) illumination include Dark Flash Photography [1], invisible marker tracking [3] and stereo depth estimation¹. In addition, the NIR band has several favourable properties that make it interesting for passive sensing. For example, dehazing [2] using NIR is effective as the longer wavelength of NIR makes it less susceptible to Rayleigh scattering. As the pixel density of CMOS and CCD sensors continues to increase, devoting some of these pixels to NIR might become an attractive proposition [4].

In this work we revisit the idea of interest point matching assuming a joint RGB+NIR capable imager. Multispectral corner detectors have been proposed previously [5, 6]. Our work is closest to [6], who suggest generalising the Harris detector by summing autocorrelation matrices per band. We show that this can be derived by consideration of the multispectral s.s.d. function, and that this formulation leads to improved matching performance for RGB+NIR registration. We also generalise the Laplacian detector [7] and perform experiments for this case as well. Colour descriptors have also been proposed, e.g., [8], where the authors investigate the effect of colour transformations on SIFT descriptors [7]. In this work,



Fig. 1. Near Infrared (NIR) images (left) often have more contrast than RGB images (right). By combining RGB and NIR channels we can improve interest point matching and image registration performance.



Fig. 2. Example of registered images from our test dataset, captured in NIR (top) and RGB (bottom).

we propose a SIFT variant specific to RGB to NIR matching. Our new descriptor will be sensitive to the sometimes large reflectance differences between the bands. We apply our new interest points and descriptors to joint RGB-NIR registration problems (see Figures 1 and 2).

In related literature, researchers have studied the statistics of images in the *far*-infrared (4-12 μ m) [9], as well as demonstrating enhanced pedestrian tracking using this band [10]. These applications require a specialist detector that is dedicated for use in the far-infrared band. In this work we focus instead on the *near*-infrared (750-1100nm), that can be captured using an ordinary digital camera.

The remainder of the paper is organised as follows: In Sections 2 and 3 we derive multispectral versions of the Harris and Laplacian detectors. Section 4 introduces the problem

¹Primesense Sensor (Microsoft Kinect): <http://www.primesense.com>

of gradient reversal and proposes a modification to SIFT that is invariant to this change. We describe our capture setup and test dataset in Section 5, and experimental results in Section 6. We close with conclusions and ideas for future work.

2. MULTISPECTRAL HARRIS

The Harris corner detector [11] finds points that maximise an s.s.d. function for any displacement $\Delta\mathbf{x}$ of a local patch:

$$e_H(\Delta\mathbf{x}) = \sum_{\mathbf{x}_i \in \mathcal{R}} w(\mathbf{x}_i) [I(\mathbf{x}_i) - I(\mathbf{x}_i + \Delta\mathbf{x})]^2, \quad (1)$$

where $I(\mathbf{x}_i)$ is the intensity at location \mathbf{x}_i , and the difference is summed with weights $w(\mathbf{x}_i)$ over region \mathcal{R} . This criterion encapsulates the idea that a corner should have distinctive 2-dimensional structure – i.e., $e_H(\Delta\mathbf{x})$ should be large for any displacement $\Delta\mathbf{x}$. We can generalise this idea to multiple spectral bands by summing the squared errors for each band

$$e_{MH}(\Delta\mathbf{x}) = \sum_{\mathbf{x}_i \in \mathcal{R}} w(\mathbf{x}_i) \sum_b [I^{(b)}(\mathbf{x}_i) - I^{(b)}(\mathbf{x}_i + \Delta\mathbf{x})]^2, \quad (2)$$

where $I^{(b)}(\mathbf{x}_i)$ denotes the intensity in band b . By Taylor expansion and dropping terms of 2nd order and higher we have

$$e_{MH}(\Delta\mathbf{x}) = \sum_{\mathbf{x}_i \in \mathcal{R}} w(\mathbf{x}_i) \sum_b [\Delta\mathbf{x}^T \nabla I^{(b)}(\mathbf{x}_i)]^2, \quad (3)$$

where $\nabla I^{(b)}$ is the 2-dimensional gradient in band b . This can be written as

$$e_{MH}(\Delta\mathbf{x}) = \Delta\mathbf{x}^T \left[\sum_b \sum_{\mathbf{x}_i \in \mathcal{R}} w(\mathbf{x}_i) \nabla I^{(b)} \nabla I^{(b)T} \right] \Delta\mathbf{x} \quad (4)$$

where

$$\mathbf{A}_M = \sum_b \sum_{\mathbf{x}_i \in \mathcal{R}} w(\mathbf{x}_i) \nabla I^{(b)} \nabla I^{(b)T} \quad (5)$$

is a 2×2 matrix which we call the multispectral autocorrelation matrix. Corners are detected where both eigenvalues of \mathbf{A}_m are large as in [11].

3. MULTISPECTRAL LAPLACIAN DETECTOR

Another common method for interest point detection is to find the extrema of a Laplacian function [7] of the smoothed image

$$f_L(\mathbf{x}) = \nabla^2 [g(\mathbf{x}; 0, \sigma^2) * I(\mathbf{x})]. \quad (6)$$

By linearity and introducing the difference of Gaussian (DoG) approximation for the Laplacian we have

$$f_D(\mathbf{x}) = [g(\mathbf{x}; 0, k\sigma^2) - g(\mathbf{x}; 0, \sigma^2)] * I(\mathbf{x}). \quad (7)$$

To extend this to a multispectral input, we take the norm of a vector of responses

$$f_{MD} = \|[g(\mathbf{x}; -, k\sigma^2) - g(\mathbf{x}; 0, \sigma^2)] * \mathbf{I}(\mathbf{x})\|_2, \quad (8)$$

where $\mathbf{I} = [I^{(0)}, I^{(1)}, \dots, I^{(n_b)}]$ is the multibanded image. By taking the vector norm instead of the sum, we prevent opposite signed gradients from cancelling each other out, as described in the next section.

4. GRADIENT DIRECTION INVARIANT SIFT

When comparing NIR and luminance (grayscale) images, one can often observe a reversal in the gradient direction, due to the differing material reflectance in RGB and NIR (see Figures 4 and 3). This poses a problem for local feature matching methods such as SIFT, which form descriptors based on a histogram of gradients. For example, the i th orientation histogram in SIFT is given by

$$h_i = \sum_{k=1}^N \left[\frac{i}{N} < \frac{\theta_k}{2\pi} < \frac{i+1}{N} \right] r_k, \quad (9)$$

where θ_k is the orientation of the k th gradient, r_k it's magnitude, and the bracketed inequality =1 if true and 0 otherwise. To achieve invariance to the gradient orientation, we form a half-size descriptor with orientations θ_k and $\theta_k + 180^\circ$ equalised, i.e.,

$$h'_i = \sum_{k=1}^{(N-1)/2} \left[\frac{i}{N} < \frac{\theta'_k}{2\pi} < \frac{i+1}{N} \right] r_k, \quad (10)$$

where

$$\theta'_k = \begin{cases} \theta_k, & 0 < \theta_k < \pi \\ \theta_k - \pi, & \pi < \theta_k < 2\pi \end{cases}. \quad (11)$$

We name the resulting descriptors **GDISIFT** (gradient-direction invariant SIFT), since they are invariant to 180° change of the gradient direction.

5. RGB-NIR IMAGE CAPTURE

Though joint RGB+NIR cameras using beam splitters exist [3], no portable versions are yet commercially available. Hence to capture a joint RGB+NIR image we take two separate exposures, using RGB and NIR filters that pass frequencies below and above 750nm, respectively. We focus on



Fig. 4. Gradient reversal: differing material reflectances can reverse the direction of gradients by 180° between RGB and NIR.



Fig. 3. Location of gradient reversals

static scenes to minimise motion of objects in the scene, and compensate for small motions of the tripod by registering the images using a feature based alignment algorithm[12]. We collected a dataset consisting of 4 scenes with images shot in RGB+NIR, and a panoramic sequence of 3-5 image pairs captured for each scene.

6. EXPERIMENTS

We perform two sets of experiments: Firstly, we investigate multispectral interest point detection in comparison to single band versions for Harris and DoG. Secondly, we test gradient-direction invariant SIFT using difference of Gaussian interest points.

We establish a ground truth registration using robust feature based registration (RANSAC [13]). We test interest point detection in 4 cases for each design: 1) luminance (grayscale) to luminance (L-L), 2) luminance to NIR (L-NIR), 3) NIR to NIR (NIR-NIR), 4) multiband to multiband (Multispectral). For each interest point design we select a fixed number of detections per image, sorted by scale from coarse to fine. For the given set of detections, we extract SIFT descriptors. We then find the closest descriptor match in each neighbouring image, and count the number of correct matches that agree with the ground truth within a tolerance $\epsilon = 3.0$ pixels. We then calculate the ratio of correct matches to the total number of matches in the area of overlap between the image – the “correct match rate”, and average this value over all image pairs.

Results are shown in Figures 5 and 6. The shaded areas show one standard deviation of the recognition rate computed over all matching image pairs. In both cases the multispec-

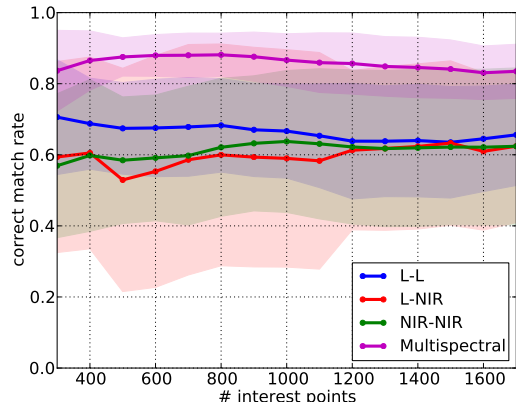


Fig. 5. Harris vs multispectral Harris.

tral interest points give significantly better performance than matching images using grayscale only. Matching NIR-NIR gave similar or worse results to L-L, so it seems to be the combination that is important. Results showing average performance over all numbers of interest points are shown in Figure 7. As expected, matching L-NIR gave the poorest results in general, with multispectral Harris and DoG giving significantly better results in both cases.

For the second experiment we tested the performance of gradient-direction invariant SIFT (GDISIFT) to ordinary SIFT when matching L-NIR. We used difference of Gaussian interest points in this test. The results are shown in Figure 8. GDISIFT gave significantly better performance as the number of interest points was increased. Note that these results are somewhat data dependent, as some images exhibit more gradient reversals than others. For example, we see that in Figure 3, most of the gradient reversals appear around the foliage, and the man-made portions maintain the same gradient direction in L and NIR.

7. CONCLUSION

We have proposed generalisations of Harris and difference of Gaussian interest points for colour plus near-infrared images, and shown that these designs give significantly improved performance in terms of matching success rate in a registration problem. We also propose a modification to the SIFT descriptor to introduce invariance to the gradient direction, which gives improved results for matching L and NIR images. Fu-

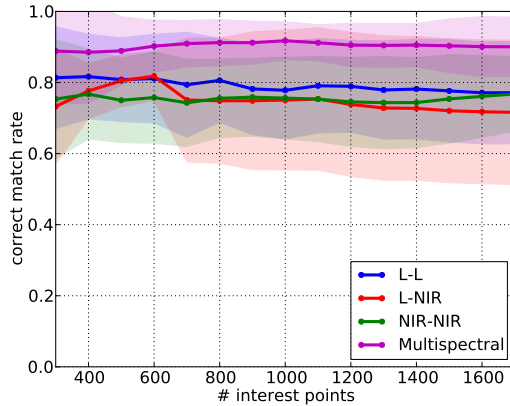


Fig. 6. DoG vs multispectral DoG.

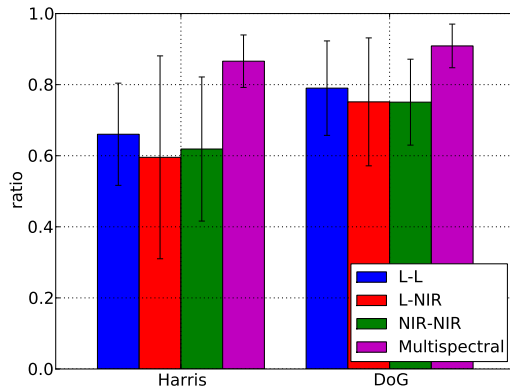


Fig. 7. Standard vs multispectral interest points.

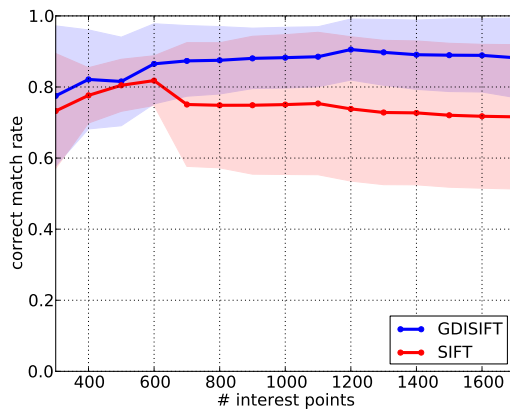


Fig. 8. SIFT vs GDISIFT

ture possibilities include investigation of multispectral region detectors, such as MSERs [14].

8. ACKNOWLEDGEMENTS

This work is supported by the National Competence Center in Research on Mobile Information and Communication Systems (NCCR-MICS), a center supported by the Swiss National Science Foundation under grant number 5005-67322, and Xerox Foundation.

9. REFERENCES

- [1] D. Krishnan and R. Fergus, "Dark flash photography," *ACM Transactions on Graphics, SIGGRAPH 2009 Conference Proceedings*, 2009.
- [2] L. Schaul, C. Fredembach, and S. Süssstrunk, "Color image dehazing using the near-infrared," in *IEEE International Conference on Image Processing*, 2009.
- [3] H. Park and J. Park, "Invisible marker tracking for AR," in *IEEE and ACM International Symposium on Mixed and Augmented Reality*, 2004.
- [4] Y. Lu, C. Fredembach, M. Vetterli, and S. Süssstrunk, "Designing color filter arrays for the joint capture of visible and near-infrared images," in *IEEE International Conference on Image Processing*, Cairo, November 2009.
- [5] C. Achard, E. Bigorgne, and J. Devars, "A sub-pixel and multispectral corner detector," in *International Conference on Pattern Recognition*, Barcelona, September, vol. 3, pp. 959–962.
- [6] J. Stöttinger, A. Hanbury, T. Gevers, and N. Sebe, "Lonely but attractive: Sparse color salient points for object retrieval and categorization," in *CVPR Workshop on Feature Detectors and Descriptors*, 2009.
- [7] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [8] K. van de Sande, T. Gevers, and C. Snoek, "Evaluating color descriptors for object and scene recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1582–1596, 2010.
- [9] N. Morris, S. Avidan, W. Matusik, and H. Pfister, "Statistics of infrared images," in *International Conference on Computer Vision and Pattern Recognition*, July 2007.
- [10] L. Zhang, B. Wu, and R. Nevatia, "Pedestrian detection in infrared images based on local shape features," in *International Conference on Computer Vision and Pattern Recognition*, Minneapolis, June 2007.
- [11] C. Harris, "Geometry from visual motion," in *Active Vision*, A. Blake and A. Yuille, Eds., pp. 263–284. MIT Press, 1992.
- [12] R. Szeliski, "Image alignment and stitching: A tutorial," *Foundations and Trends in Computer Graphics and Computer Vision*, vol. 2, no. 1, pp. 1–104, December 2006.
- [13] M. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with application to image analysis and automated cartography," *Communications of the ACM*, vol. 24, pp. 381–395, 1981.
- [14] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in *British Machine Vision Conference (BMVC'02)*, 2002.