DASC: Dense Adaptive Self-Correlation for Multi-modal and Multi-spectral Correspondence

Seungryong Kim, Dongbo Min, Bumsub Ham, Seungchul Ryu, Minh N. Do, and Kwanghoon Sohn

YONSEI UNIVERSITY · CNU 충남대학교 · Inria INVENTEURS DU MONDE NUMÉRIQUE · ILLINOIS UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

IEEE 2015 Conference on Computer Vision and Pattern Recognition
CVPR 2015 BOSTON JUNE 7-12

Introduction

❖ Motivation

- In multi-modal and multi-spectral image, conventional descriptors often fail to estimate correspondence

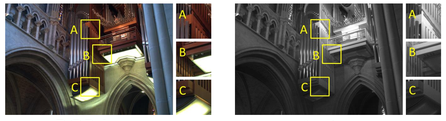
❖ Goal

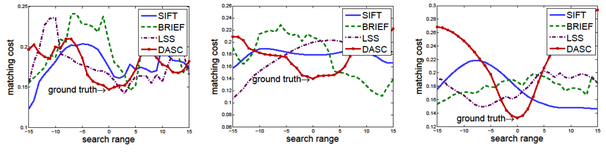
- To establish dense correspondence for those images

Background

❖ Challenge on Multi-modal Images

- Nonlinear photometric deformation, e.g., gradient reverses and intensity order variations





❖ Limitation of Conventional Descriptors

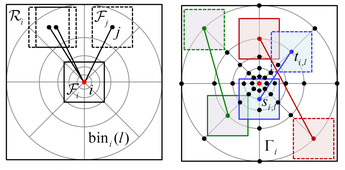
- Image gradient (SIFT) or intensity comparison (BRIEF) cannot capture coherent matching evidence

DASC Descriptor

❖ Randomized Receptive Field Pooling

- Unlike center-biased max pooling in LSS descriptor, the DASC descriptor incorporates randomized receptive field pooling

- **Intuitions**
  - ✓ Center-biased pooling sensitive to noise
  - ✓ Randomness of BRIEF

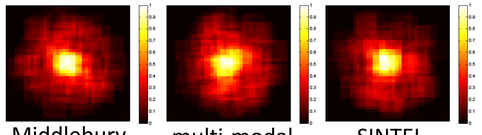


- ✓ LSS descriptor $d_{i,l}^{\mathrm{LSS}} = \max_{j \in \mathrm{bin}_i(l)} \mathcal{C}(i,j)$
- ✓ DASC descriptor $d_{i,l} = \mathcal{C}(s_{i,l}, t_{i,l}), \quad s_{i,l}, t_{i,l} \in \Gamma_i$

❖ Sampling Pattern Learning

- Exploit supports vector machine (SVM) with linear kernel
- Features: $r_{m,l} = \exp\left(-(d_{m,l}^1 - d_{m,l}^2)^2 / 2\sigma_r^2\right)$
- Energy function for SVM

$$\mathcal{L}(v) = \lambda \| v \|^2 + \sum_m \max(0, 1 - y_m \rho(r_m))$$



❖ Small Support Window Similarity

- Adaptive self-correlation measure

$$\Psi(s,t) = \frac{\sum_{s',t'} \omega_{s,s'} \omega_{t,t'} (f_{s'} - \mathcal{G}_s)(f_{t'} - \mathcal{G}_t)}{\sqrt{\sum_{s'} \{\omega_{s,s'} (f_{s'} - \mathcal{G}_s)\}^2} \sqrt{\sum_{t'} \{\omega_{t,t'} (f_{t'} - \mathcal{G}_t)\}^2}}$$

- Robust estimation

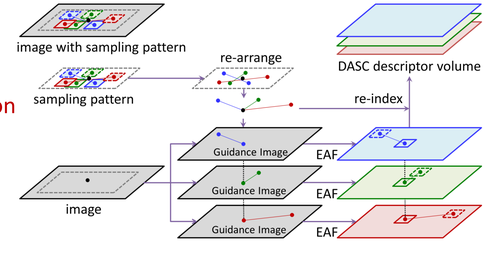
$$\mathcal{C}(s,t) = \max(\exp(-(1 - |\Psi(s,t)|) / \sigma), \tau)$$

- The correlation $\mathcal{C}(s,t)$ is normalized with norm of all l

❖ Efficient Computation

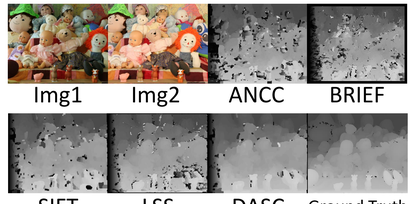
- Asymmetric weights $\omega_{s,s'}$
- Reference-biased sampling pairs $(i,j) = (i, i + t_{i,l} - s_{i,l})$
- Approximated adaptive self-correlation

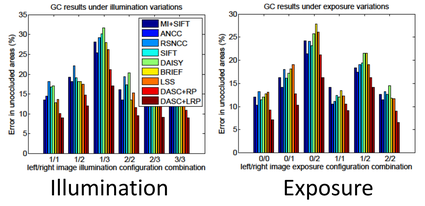
$$\tilde{\Psi}(i,j) = \frac{\sum_{i',j'} \omega_{i,i'} (f_{i'} - \mathcal{G}_i)(f_{j'} - \mathcal{G}_{i,j})}{\sqrt{\sum_{i'} \omega_{i,i'} (f_{i'} - \mathcal{G}_i)^2} \sqrt{\sum_{i',j'} \omega_{i,i'} (f_{j'} - \mathcal{G}_{i,j})^2}}$$



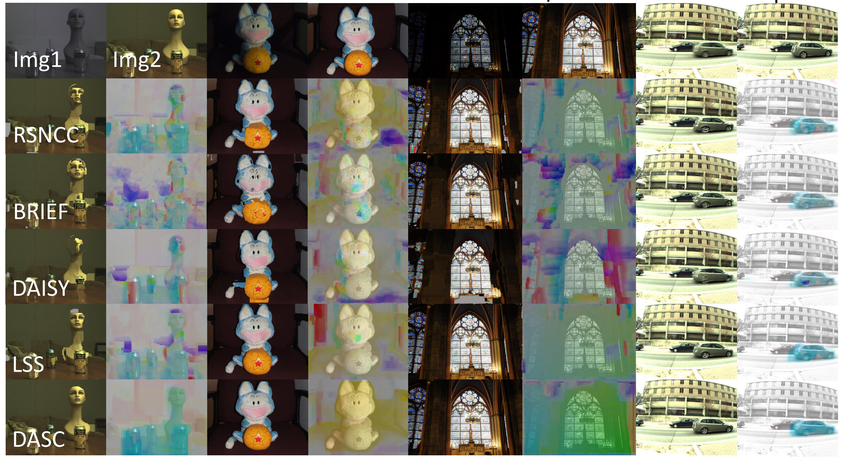
Experimental Results

❖ Middlebury Stereo Benchmark





❖ Multi-modal and Multi-spectral Image Pairs



Conclusion

- Robust novel descriptor called the DASC for multi-modal correspondence
- Leverages adaptive self-correlation and randomized receptive pooling
- Efficient computation with fast edge-aware filters

Download the code at http://seungryong.github.io/DASC/