Improving Image Matching with Varied Illumination

Sarah Braeger **Hassan Foroosh**

Computational Imaging Lab, University of Central Florida

Computational Imaging Lab. (CIL)
Department of Computer Science

Abstract

We present a method to maximize feature matching performance across stereo image pairs by varying illumination. We perform matching between views per lighting condition, finding unique SIFT correspondences for each condition. These feature matches are then collected together into a single set, selecting those features which present the highest quality match. Instead of capturing each view under each illumination, we approximate lighting changes with a pretrained relighting convolutional neural network which only requires each view captured under a single specified lighting condition. We then collect the best of these feature matches over all lighting conditions offered by the relighting network. We further present an optimization to limit the number of lighting conditions evaluated to gain a specified number of matches. Our method is evaluated on a set of indoor scenes excluded from training the network with comparison to features extracted from pretrained VGG16. Our method offers an average 5.5× improvement in number of correct matches while retaining similar precision than by the original lit image pair per scene alone.

Figure 1:The Kingston Living 5_6 scene view pair with the set of SIFT feature matches. The top row is on the input single illumination. The second row is the result of our method, those matches merged over 25 lighting conditions synthesized by relighting. Matches illustrated as lines across views.

Relighting the View Pair

We generate images with a pretrained relighting network [1]. The network takes in a single image per view A: $\{(I_0^A, L_0)\}$, where L_0 is the lighting condition in Ω for index 0. I_j^A is an image taken with that lighting condition L_j from view A. The set of images reconstructed by the network which we then use in matching for a given view A: $R^A : \{(R_j^A, L_j)|j = 1...|\Omega|\}$, where R_j^A is a reconstructed image for view A at lighting condition j and $|\Omega| = 25$, the total of lighting conditions configured for the network.

Matching across Views

We perform feature detection and matching with SIFT across a pair of views A and B for the set of relit images R^A and R^B. Matching is performed per lighting condition L_j in the $|\Omega|$ possible conditions per view pair forming a set of matches M_j. The scene-dependent individual lighting condition L_j which maximizes matching performance of number of correct matches NCM per image pair:

$$j = \underset{j \in |\Omega|}{\operatorname{argmax}}\, NCM(M_j) \tag{1}$$

Match Merging

We collect match sets M_j over all $|\Omega|$ lit images per view. We define the set of merged matches $\mathbf{M}$ as the union of sets of matches $\{M_j|\forall j \in \Omega\}$ with conflicts resolved. When a match m_1 from M_j conflicts with the same pixel coordinate $\mathbf{x}_j^A$ from a match m_2, we select the match with the highest $quality = 1 - distance$, where distance is defined according to Lowe's ratio test.The set of merged matches $\mathbf{M}$:

$$\mathbf{M} = \bigcup_{\mathbf{x}^A=0}^{N} \operatorname{argmax}_{m_{j\in|\Omega|}} quality(m_j) \tag{2}$$

where $\mathbf{x}^A$ is a pixel coordinate from view I^A and N is the total number of pixels in I^A. m_j is a match in the set of matches M_j. We select the match of maximum quality over all lighting conditions $|\Omega|$ which share a pixel coordinate in I^A.

Experimental Setup

We experiment with the recently released multi-illumination dataset [1], which has discretized lighting and perspective change in capturing views. We manually selected 37 scene pairs with sufficient overlap and static scene content. These view pairs are specified in Figure 3. We compare results to matches discovered on a single illumination with features extracted from pretrained VGG16 [2] on ImageNet via the image registration method presented in [3].
We define ground truth as those matches which satisfy the epipolar constraint

$$x'^T F x = 0 \tag{3}$$

where the fundamental matrix F is estimated from the inlier SIFT correspondences of input image capture pair for lighting condition $L0$. This criterion is selected due to the uncalibrated camera capture scenario of the multi-illumination dataset.

Precision

We select the top-k matches with a set of strategies for $L0$ and merged set $\mathbf{M}$ matches over all scenes, where $k = 100$. The precision for the top-k matches in Figure 2. There is a slight improvement in overall precision between fixing for the captures taken only with $L0$ and the matches from the full set of relit images over all lighting conditions. The most improvement is from selecting the top-k most frequently detected locations of matches in the left view. This strategy is ideal for a top-k set of matches, but merging for overall match quality is recommended for the full set of merged matches.

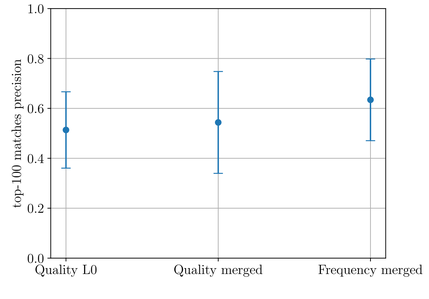

Figure 2:Precision for top-100 matches for $L0$ and merged set $\mathbf{M}$ over all lighting conditions.

Correct Matches

The number of correct matches for all scenes are plotted in Figure 3. The number of correct matches per scene is consistently higher than from the input $L0$ lit captures.

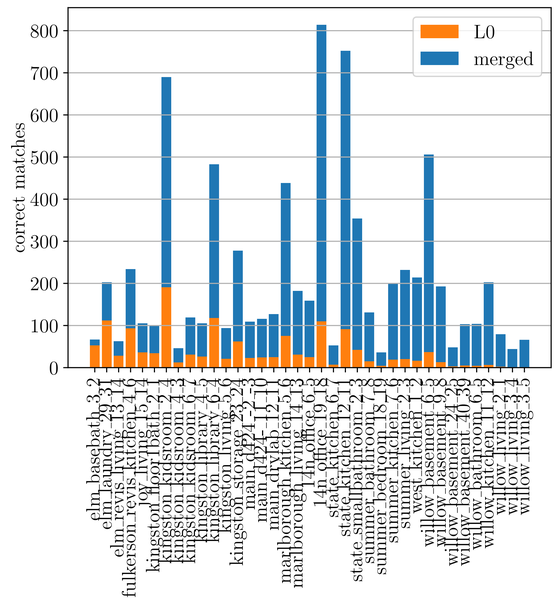

Figure 3:Correct matches for input capture pair at $L0$ and merged set $\mathbf{M}$ for all scene view pairs.

Features	Average	Std. Dev.
SIFT0	38.270	41.441
VGG16	68.216	28.268
OURS	212.189	198.612

Table 1:Number of correct matches with SIFT on input capture pair only, VGG16 features, and our merged matches.

References

[1] L. Murmann, M. Gharbi, M. Aittala, and F. Durand, "A multi-illumination dataset of indoor object appearance," in *2019 IEEE International Conference on Computer Vision (ICCV)*, Oct 2019.

[2] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations*, 2015.

[3] Z. Yang, T. Dan, and Y. Yang, "Multi-temporal remote sensing image registration using deep convolutional features," *IEEE Access*, vol. 6, pp. 38 544–38 555, 2018.