Improving Image Matching with Varied Illumination

Sarah Braeger Hassan Foroosh

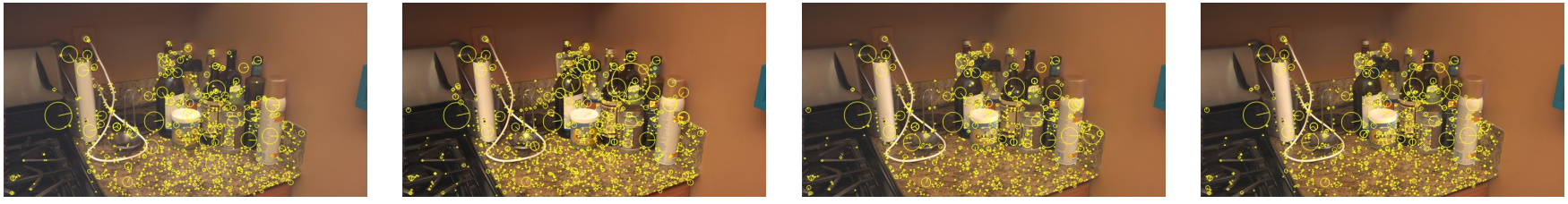
Computational Imaging Lab
University of Central Florida

Key Components

- Vary lighting to increase features.
- Don't physically capture lighting: generate images with relighting network.
- Merge features across lighting conditions.
- Increase **number of correct matches** by $5.5x$ for indoor stereo images. Retain similar precision.

Insight

- Varying lighting gives different features for matching.
- Synthesize lighting changes with a relighting network.



(a) $L0$ (b) $L2$ (c) $L5$ (d) $L6$

Figure: SIFT keypoints on relit generated images for State Kitchen 7.

Relighting the View Pair

- Generate images with a pretrained relighting network [1].
- Input single image per view A: $\{(I_0^A, L_0)\}$, where L_0 is the specified lighting condition for index 0.
- R_j^A is a reconstructed image for view A for lighting condition $L_j \in \Omega$.
- Total synthesized lighting conditions: $|\Omega| = 25$.

Matching across Views

- Match pair of views A and B for the set of relit images R^A and R^B.
- Performed **per lighting condition** L_j in the $|\Omega|$ possible, forming a set of matches M_j.

Match Merging

- The set of merged matches $\mathbf{M}$:

$$\mathbf{M} = \bigcup_{\mathbf{x}^A=0}^{N} \mathrm{argmax}_{m_{j \in |\Omega|}} \mathit{quality}(m_j) \tag{1}$$

- $\mathbf{x}^A$: pixel coordinate from view I^A.
- N: total number of pixels in I^A, m_j: a match in the set of matches M_j.

Handling Match Merging Conflicts

- When a match m_1 from M_j conflicts with the same pixel coordinate $\mathbf{x}_j^A$ from a match m_2:
- Select the match with the highest $\mathit{quality} = 1 - \mathit{distance}$, defined by Lowe's ratio test.

Experimental Setup

- We experiment with the recently released **multi-illumination dataset** [1], manually selecting pairs with sufficient stereo overlap.
- We define ground truth as those matches which satisfy the **epipolar constraint**:

$$x'^T F x = 0 \tag{2}$$

- F is estimated from the inlier SIFT correspondences of input image capture pair for lighting condition $L0$.

Comparison

Figure: The Kingston Living 5_6 scene view pair with set of SIFT matches. The **top row** is on the input single lit pair. The **bottom row** is the result of our method, those matches merged over 25 lighting conditions synthesized by relighting.

Precision

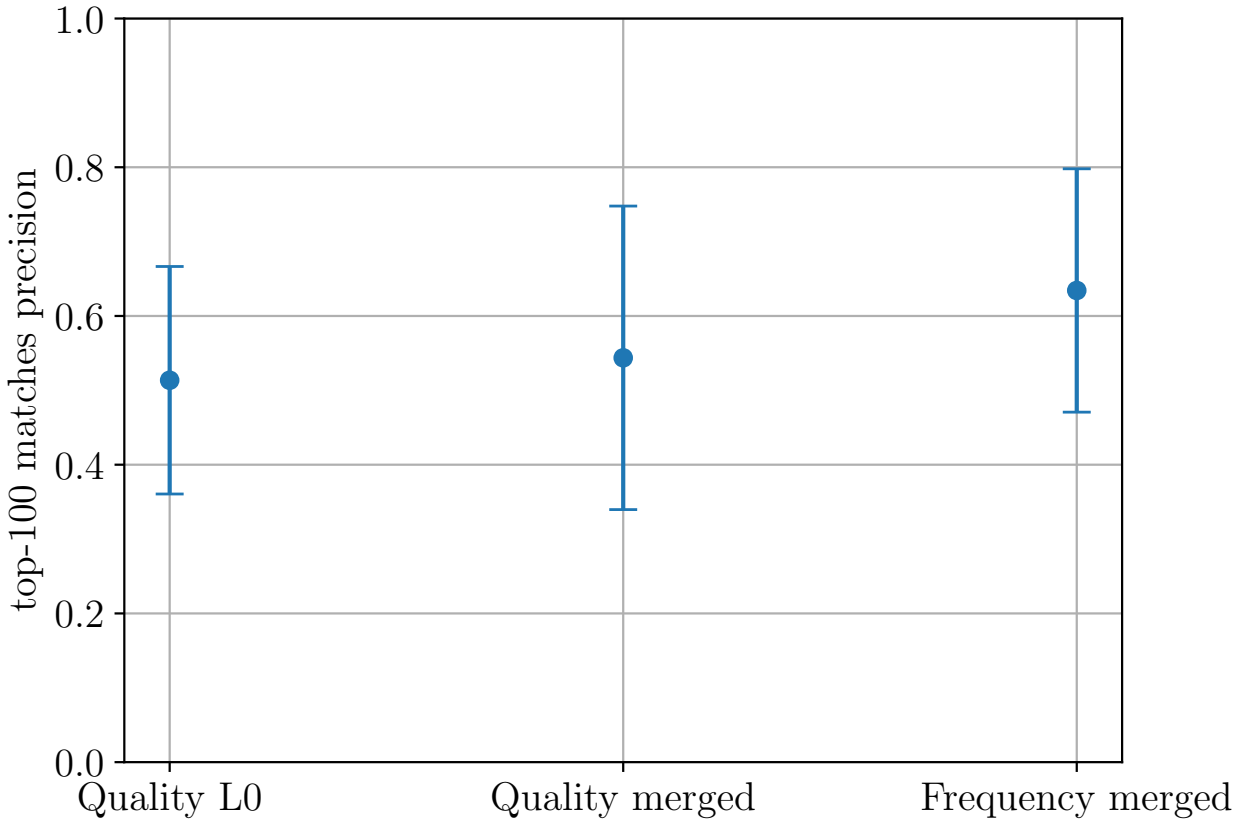


Figure: Precision for top-100 matches for *L*0 and merged set **M** over all lighting conditions.

Number of Correct Matches

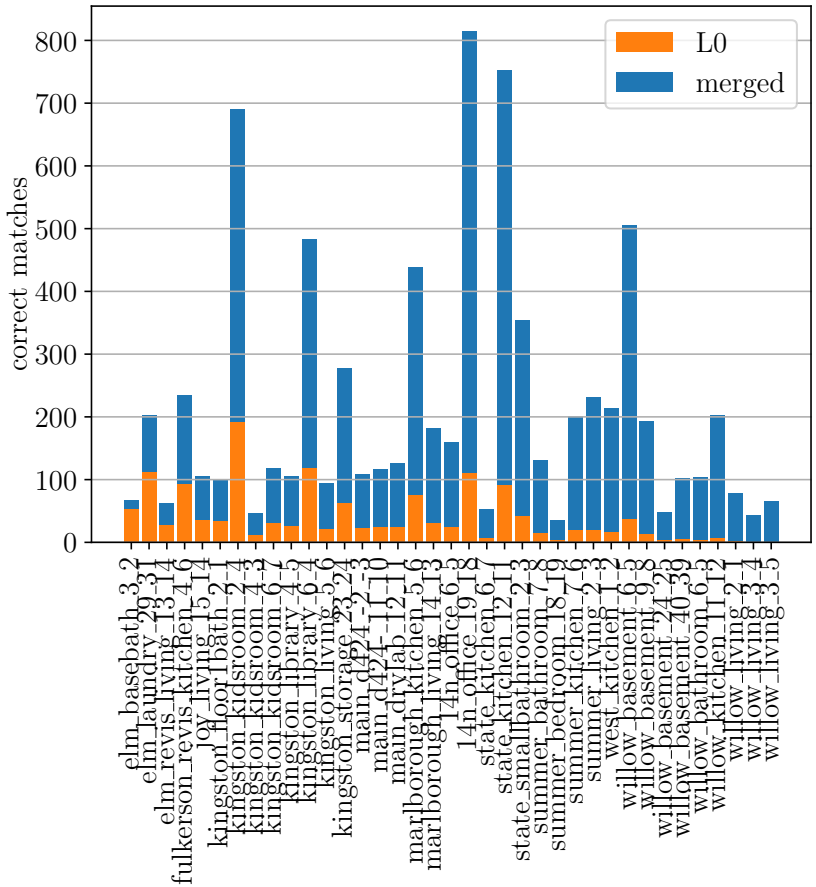


Figure: Correct matches for input capture pair at *L0* and merged set **M** for all scene view pairs.

Comparing Correct Matches

L0:

VGG16 [2]:

Ours:

Figure: The Kingston Kidsroom 6_7 scene view pair.

- Our match density is higher from nearby local gradients.

Comprehensive Comparison

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

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

References

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