Evaluating Local Features for Day-Night Matching

Hao Zhou[1], Torsten Sattler[2], David W. Jacobs[1]

[1] University of Maryland, College Park, MD, USA,
[2] Department of Computer Science, ETH Zurich, Switzerland
hzhou@cs.umd.edu, sattlert@inf.ethz.ch, djacobs@umiacs.umd.edu

Abstract. This paper evaluates the performance of local features in the presence of large illumination changes that occur between day and night. Through our evaluation, we find that repeatability of detected features, as a de facto standard measure, is not sufficient in evaluating the performance of feature detectors; we must also consider the distinctiveness of the features. Moreover, we find that feature detectors are severely affected by illumination changes between day and night and that there is great potential to improve both feature detectors and descriptors.

Keywords: Feature detector, day-night image matching, performance evaluation, illumination changes

1 Introduction

Feature detection and matching is one of the central problems in computer vision and a key step in many applications such as Structure-from-Motion [18], 3D reconstruction [3], place recognition [20], image-based localization [27], Augmented Reality and robotics [7], and image retrieval [17]. Many of these applications require robustness under changes in viewpoint. Consequently, research on feature detectors [10, 25, 12, 8, 24, 13] and descriptors [6, 2, 16, 22] has for a long time focused on improving their stability under viewpoint changes. Only recently has robustness against seasonal [19] and illumination changes [24, 21] come into focus. Especially the latter is important for large-scale localization and place recognition applications, *e.g.*, for autonomous vehicles. In these scenarios, the underlying visual representation is often obtained by taking photos during daytime and it is infeasible to capture large-scale scenes also during nighttime.

Many popular feature detectors such as Difference of Gaussians (DoG) [6], Harris-affine [9], and Maximally Stable Extremal Regions (MSER), as well as the popular SIFT descriptor [6] are invariant against (locally) uniform changes in illumination. However, the illumination changes that can be observed between day and night are often highly non-uniform, especially in urban environments (*cf.* Fig. 1). Recent work has shown that this causes problems for standard feature detectors: Verdie *et al.* [24] demonstrated that a detector specifically trained to handle temporal changes significantly outperforms traditional detectors in challenging conditions such as day-night illumination variations. Torii

et al. [20] observed that foregoing the feature detection stage and densely extracting descriptors instead results in a better matching quality when comparing daytime and nighttime images. Naturally, these results lead to a set of interesting questions: (i) to what extent is the feature detection stage affected by the illumination changes between day and night? (ii) the number of repeatable features provides an upper bound on how many correspondences can be found via descriptor matching. How tight is this bound, *i.e.*, is finding repeatable feature detections the main challenge of day-night matching? (iii) how much potential is there to improve the matching performance of local detectors and descriptors, *i.e.*, is it worthwhile to invest more time in the day-night matching problem?

In this paper, we aim at answering these questions through extensive quantitative experiments, with the goal of stimulating further research on the topic of day-night feature matching. We are interested in analyzing the impact of day-night changes on feature detection and matching performance. Thus, we eliminate the impact of viewpoint changes by collecting a large dataset of daytime and nighttime images from publicly available webcams [5] [1]. Through our experiments on this large dataset, we find that: (i) the repeatability of feature detectors for day-night image pairs is much smaller than that for day-day and night-night image pairs, meaning that detectors are severely affected by illumination changes between day and night; (ii) for day-night image pairs, high repeatability of feature detectors does not necessarily lead to a high matching performance. For example, the TILDE [24] detector specifically learned for handling illumination changes has a very high repeatability, but the precision and recall of matching local features are very low. A low recall shows that the number of repeatable points provides a loose bound for the number of correspondences that could be found via descriptor matching. As a result, further research is necessary for improving both detectors and descriptors; (iii) through dense local feature matching, we find that there are a lot more correspondences that could be found using local descriptors than are produced by current detectors, *i.e.*, there is great potential to improve detectors for day-night feature matching.

2   Dataset

Illumination and viewpoint changes are two main factors that would affect the performance of feature detectors and descriptors. Ideally, both detectors and descriptors should be robust to both type of changes. However, obtaining a large dataset with both types of changes with ground truth transformations is difficult. In this paper, we thus focus on pure illumination changes and collect data that does not contain any viewpoint changes. Our results show that already this simpler version of the day-night matching problem is very hard.

The AMOS dataset [5], which contains a huge number of images taken (usually) every half an hour by outdoor webcams with fixed positions and orientations, satisfies our requirements perfectly. [24] has collected 6 sequences of

---

[1] Please find the data set at http://www.umiacs.umd.edu/ hzhou/dnim.html

Fig. 1. Images taken from 00:00 - 23:00 in one image sequence of our dataset.

images taken at different times of the day for training illumination robust detectors from the AMOS dataset. However, the dataset has no time stamps and some of the sequences have no nighttime images. As a consequence, we collect our own dataset from AMOS. 17 image sequences with relatively high resolution containing 1722 images are selected. Since the time stamps of the images provided by AMOS are usually not correct, we choose image sequences with time stamp watermarks. The time of the images will be decided by the watermarks which are removed afterwards. For each image sequence, images taken in one or two days are collected. Fig. 1 gives an example of images we collected.

3 Evaluation

3.1 Keypoint Detectors

For evaluation, we focus on the keypoint detectors most commonly used in practice. We choose DoG [6], Hessian, HessianLaplace, MultiscaleHessian, Harris-Laplace and MultiscaleHarris [9] implemented by vlfeat [23] for evaluation. Their default parameters are used to determine how well these commonly used settings perform under strong illumination changes. DoG detects feature points as the extrema of the difference of Gaussian functions. By considering the extrema of the difference of two images, DoG detections are invariant against additive or multiplicative (affine) changes in illumination. Hessian, HessianLaplace and MultiscaleHessian are based on the Hessian matrix $\begin{pmatrix} L_{xx}(\sigma) & L_{xy}(\sigma) \\ L_{yx}(\sigma) & L_{yy}(\sigma) \end{pmatrix}$, where L represents the image smoothed by a Gaussian with standard deviation σ and L_{xx}, L_{xy}, and L_{yy} are the second-order derivatives of L. Hessian detects feature points as the local maxima of the determinant of the Hessian matrix. Hessian-Laplace chooses a scale for the Hessian detector that maximizes the normalized Laplacian $|\sigma^2(L_{xx}(\sigma) + L_{yy}(\sigma))|$. MultiscaleHessian instead applies the Hessian detector on multiple scales of images and detects feature points at each scale independently. HarrisLaplace and MultiscaleHarris extended the Harris corner detector to multiple scales in a similar way to HessianLaplace and Multiscale-Hessian. The Harris corner detector is based on the determinant and trace of the
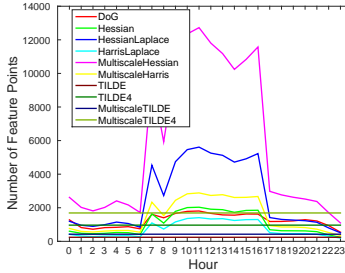
Fig. 2. The number of feature points detected at different time

second moment matrix of gradient distribution. All these gradient based methods are essentially invariant to additive and multiplicative illumination changes.

We also included the learning based detector TILDE [24], since it is designed to be robust to illumination changes. We use the model trained on the St. Louis sequence as it has the highest repeatability when testing on the other image sequences [24]. TILDE detects feature points at a fixed scale. In this paper, we define a multiple scale version by detecting features at multiple scales, denoted as MultiscaleTILDE. Feature points are detected from the original image and images smoothed by a Gaussian with standard deviation of 2 and 4. When TILDE detects feature points from the original image, the scale of it is set to be 10. Accordingly, the scale of detected feature points from those three images are set to be 10, 20 and 40. As suggested by [24], we keep a fixed number of feature points based on the resolution of the image. For the proposed MultiscaleTILDE, the same number of feature points as that of TILDE are selected for the first scale. For other scales, the number of feature points selected are reduced by half compared with the previous scale. In modified versions, we include 4 times as many points as suggested, naming these TILDE4 and MultiscaleTILDE4 respectively.

3.2   Repeatability of Detectors

In this section we address the question: **to what extent are feature detections affected by illumination changes between day and night?** by evaluating how many feature points are detected, and how repeatable they are. First we show the number of detected feature points at different times of the day for different detectors in Fig. 2. The numbers are averaged from all 17 image sequences in our dataset. The number of feature points for TILDE is the same across different times, since a fixed number of feature points are extracted. For the other detectors, fewer feature points are detected at nighttime. Especially, the number of feature points detected by HessianLaplace and MultiscaleHessian are affected most by illumination changes between day and night.

We then use the repeatability of the detected feature points to evaluate the performance of detectors. According to [10], the measurement of repeatability is related to the detected region of feature points. Suppose σ_a and σ_b are the scale of two points A and B, (x_a, y_a) and (x_b, y_b) are their locations, the detected

regions μ_a and μ_b are defined as the region of $(x - x_a)^2 + (y - y_a)^2 = (3\sigma_a)^2$ and $(x - x_b)^2 + (y - y_b)^2 = (3\sigma_b)^2$ respectively, where 3σ is the size of one spatial bin from which the SIFT feature is extracted. Then A and B are considered to correspond to each other if $1 - \frac{\mu_a \cap \mu_b}{\mu_a \cup \mu_b} \leq 0.5$, *i.e.* the intersection of these two regions are larger than or equal to half of the union of these two regions. This overlap error is the same as the one proposed in [10] except that we do not normalize the region size. This is because if the detected regions do not overlap, we cannot extract matchable feature descriptors; normalizing the size of the region would obscure this. For example, two regions with small scales may be judged to correspond after normalization. However, the detected region from which the feature descriptor is extracted may not overlap at all, making it impossible to extract feature descriptors to match them.

Some of the images in our dataset may contain moving objects. To avoid the effect of those objects, we define "ground truth" points and compute the repeatability of detectors at different times w.r.t. them. To make the experiments comprehensive, we use daytime ground truth and nighttime ground truth. Images taken at 10:00 to 14:00 are used to get the daytime ground truth feature points (and 00:00 to 02:00 together with 21:00 to 23:00 for nighttime ground truth feature points). We select the image that has the largest number of detected feature points and match them to those in other images in that time period. A feature point is chosen as a ground truth if it appears in more than half of all the images of that time period. Fig. 3 (a) and (b) shows the number of daytime and nighttime ground truth feature points detected for different detectors respectively. We notice that though Fig. 2 shows the number of detected feature points for TILDE4 at daytime is the second smallest among all the detectors, the number of daytime ground truth feature points of TILDE4 is larger than 6 detectors. This implies that the feature points detected by TILDE4 for daytime images are quite stable across different images.

We use these ground truth feature points to compute the repeatability of the chosen detectors over different times of the day. Thus, repeatability is determined by measuring how often the ground truth points are re-detected. Fig. 3 (c) and (d) show that the repeatability of features for nighttime images w.r.t. nighttime ground truth is very high for all the detectors; this is because the illumination of nighttime images are usually quite stable without the effect of sunlight (*cf.* Fig. 1). For comparison, the repeatability of daytime images w.r.t. daytime ground truth is smaller and the performance of different detectors varies a lot. Moreover, both Fig. 3 (c) and (d) show that the repeatability of day-night image pairs is very low for most detectors, which implies that detectors are heavily affected by day-night illumination changes. The drop-off between 05:00-07:00 and 17:00-18:00 is caused by illumination changes between dusk and dawn. The peaks of the repeatability, as 09:00 in Fig. 3 (c) and 03:00 and 20:00 in Fig. 3 (d), appear because they are close to the time from which the ground truth feature points are computed. Among all the detectors, both single scale and multiple scale TILDE have high repeatabilities of around 50% for day-night image pairs. This is not surprising since the TILDE detector was constructed to be robust to
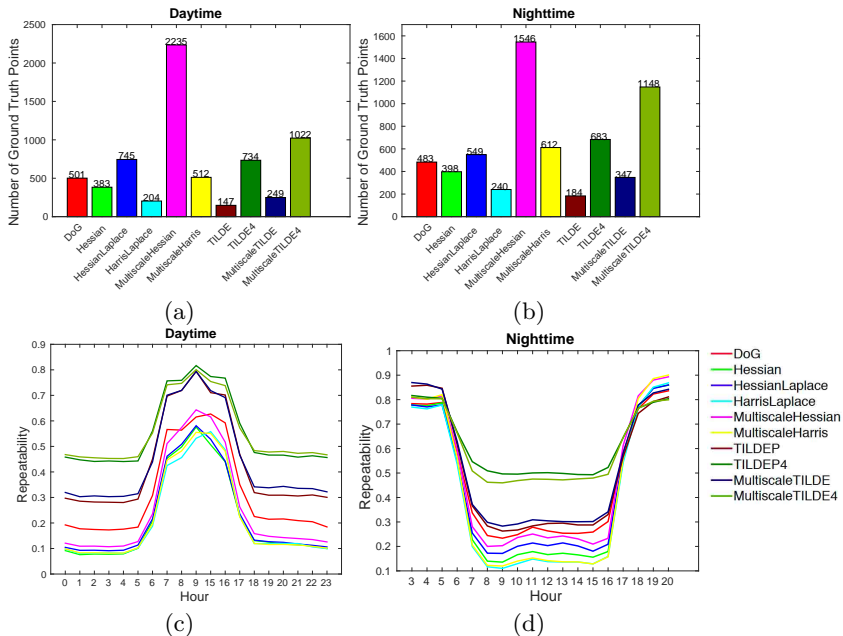
Fig. 3. (a) and (b) show average number of daytime and nighttime ground truth feature points for each detector respectively. We show repeatability of different detectors at different times of the day w.r.t. (c) daytime and (d) nightime ground truth feature points. Please note time periods that are used to compute the ground truth feature points are excluded for fair comparison.

illumination changes by learning the impact of these changes from data. Based on the fact that nearly every second TILDE keypoint is repeatable, we would expect that TILDE is well-suited for feature matching between day and night.

3.3   Matching Day-Night Image Pairs

In theory, every repeatable keypoint should be matchable with a descriptor since its corresponding regions in the two images have a high overlap. In practice, the number of repeatable keypoints is only an upper bound since the descriptors extracted from the regions might not match. For example, local illumination changes might lead to very different descriptors. In this section, we thus study the performance of detector+descriptor on matching day-night image pairs. We try to answer the question **whether finding repeatable feature detections is the main challenge of day-night feature matching**, *i.e.*, whether finding repeatable keypoints is the main bottleneck or whether additional problems are created by the descriptor matching stage. We use both precision and recall of feature descriptor matching to answer this question. Suppose for a day-night image pair, N true matches are provided by detectors. N_f matched feature points are found by matching descriptors, among which N_c matches are true matches.
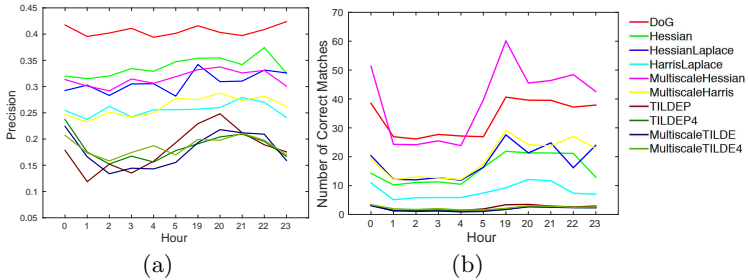
Fig. 4. (a) shows the precision of RootSIFT matching of day-night image pairs for different detectors at different nighttime. (b) shows their corresponding number of correct matched feature points.

Then the precision and recall of detector+descriptor are defined as N_c/N_f and N_c/N, respectively. Precision is a usual way to evaluate the accuracy of matching by detector+descriptor. Recall, on the other hand, tells us what is the main challenge to increase the number of matches. A low recall means improving feature descriptors is the key to getting more matches. On the contrary, feature detection is the bottleneck for getting more matches if a high recall is observed, but still an insufficient number of matching features are found.

For each image sequence, images taken at 00:00 - 05:00 and 19:00 - 23:00 are used as nighttime images and those taken at 09:00 - 16:00 are daytime images. One image is randomly selected from each hour in these time periods and every nighttime image is paired with every daytime image to create the day-night image pairs. As the SIFT descriptor is still the first choice in many computer vision problems, and its extension, RootSIFT [1] performs better than SIFT, we use RootSIFT as the feature descriptor. To match descriptors, we use nearest neighbor search and apply Lowe's ratio test [6] to remove unstable matches. The default ratio provided by vlfeat [23] is used in our evaluation. In practice, the ratio test is used to reject wrong correspondences but also rejects correct matches. The run-time of subsequent geometric estimation stages typically depends on the percentage of wrong matches. Sacrificing recall for precision is thus often preferred since there is enough redundancy in the matches.

The precisions of matching day-night images for all the detectors at different daytimes are shown in Fig. 4 (a). We found that though different versions of TILDE have the highest repeatability among all the detectors, in general, they have the lowest precision. There are more than 20% drop in precision w.r.t. DoG in most cases. This shows that a higher repeatability of a detector does not necessarily mean better performance for finding correspondences, and detectors and descriptors are highly correlated with each other for matching feature points. As shown in Fig. 4 (b), the number of correct matches detected by RootSIFT for all the detectors are quite small. Even for detectors like DoG, which has the highest precision, only 20 - 40 correct matches can be found. As a result,
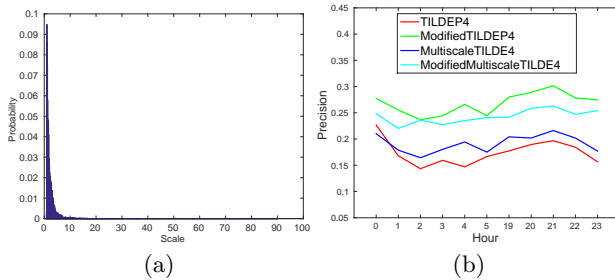
Fig. 5. (a) shows the histogram of scales for correctly matched RootSIFT features using DoG as detector. (b) compares the precision of TILDE4 and MultiscaleTILDE4 with small and large scales.
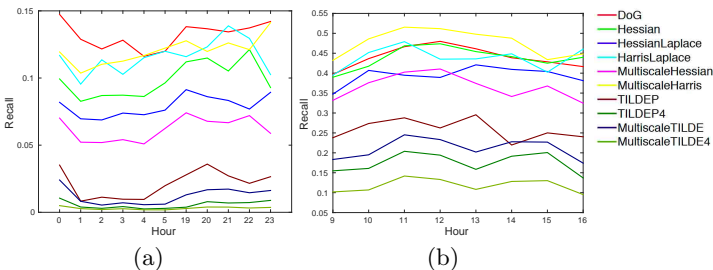

Fig. 6. (a) shows the recall of RootSIFT matching of day-night image pairs for different detectors at different nighttimes. (b) shows the recall of RootSIFT matching of day-day image pairs for different detectors at different daytimes.

for applications that need a large number of matches between day-night image pairs, the performance of these detector+descriptor may not be satisfactory.

Another interesting finding from Fig. 4 is that the multiple scale version of TILDE, MultiscaleTILDE4, does not have a higher precision compared with TILDE4. One possible reason may be that the scales of features detected by MultiscaleTILDE4 are set to 10, 20 and 40, which are too large. Fig. 5 (a) shows that most of the correctly matched features of DoG+RootSIFT have scales within 10. This is because within a smaller region, the illumination changes between day and night images are more likely to be uniform, to which the RootSIFT feature is designed to be robust. To better understand the effect of scales, we set the scale of TILDE4 to be 1 and scales of MultiscaleTILDE4 to be 1, 2 and 4, and denote these two modified versions as ModifiedTILDE4 and Modified-MultiscaleTILDE4. Fig. 5 (b) compares the precision of them with TILDE4 and MultiscaleTILDE4. We find that by setting the scale to be small, there is around $5\% - 10\%$ increase in precision. Intuitively, with larger scales, features may contain more information, which should be beneficial for matching. However, descriptors will need to be trained to make good use of those information, especially when robustness under viewpoint changes is required.
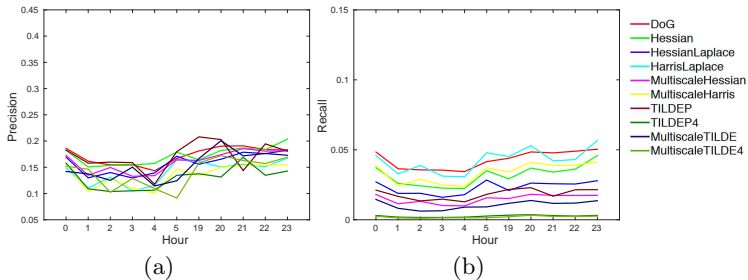
(a)   (b)

Fig. 7. (a) shows the precision of matching of day-night image pairs for different detectors at different nighttime using cnn feature. (b) shows the recall of matching of day-night image pairs for different detectors at different nighttime using cnn feature.

To examine the main challenge of finding more correspondences, the recall of these detectors for matching day-night image pairs is shown in Fig. 6 (a). We find that the recall of each detector is very low. As a consequence, **one way to improve the performance of day-night image matching is to improve the robustness of descriptors to severe illumination changes**. As shown in Fig. 6, a much higher recall can be noticed for day-day image pairs, meaning that RootSIFT is robust to small illumination changes at daytime. However, it is not so robust to severe illumination changes between day and night. The low recall of day-night image pairs implies that there are a lot of "hard" patches from which RootSIFT cannot extract good descriptors.

With the development of deep learning, novel feature descriptors based on convolutional neural networks have been proposed. Many of them [15, 4, 26] outperform SIFT. We choose the feature descriptor proposed in [15] as an example to evaluate the performance of the learned descriptor+detector. [15] is chosen since their evaluation method is Euclidean distance, which can be used easily in our evaluation framework. Fig. 7 shows that this CNN feature performs even worse than RootSIFT+detectors. One reason is that [15] is learned from the data provided by [2], which mainly focus on viewpoint changes, and illumination changes that are small. Though [15] shows its robustness to small illumination changes as in DaLI dataset [14], it is not very robust to illumination changes between day and night in our dataset.[2]

4 Potential of Improving Detectors

In this section, we try to examine **the potential of improving feature detectors** by fixing the descriptor to be RootSIFT.

Inspired by [20], we extract dense RootSIFT features from day-night image pairs for matching. When doing the ratio test, we select the neighbor that

---

[2] We also tried to tune the descriptor using day-night patch pairs, but were not able to increase the descriptor's performance.

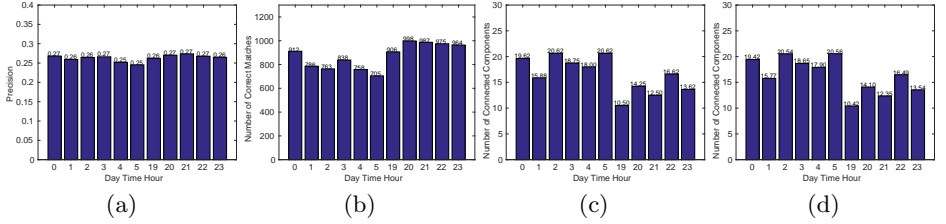(a)   (b)   (c)   (d)

Fig. 8. (a) shows the precision of matching dense RootSIFT for day-night image pairs at different nighttimes. (b) shows the number of correct matches of dense RootSIFT for day-night image pairs. (c) shows the number of connected components of matched points at different nighttime. (d) shows the number of connected components that contain no matched points of DoG+RootSIFT.
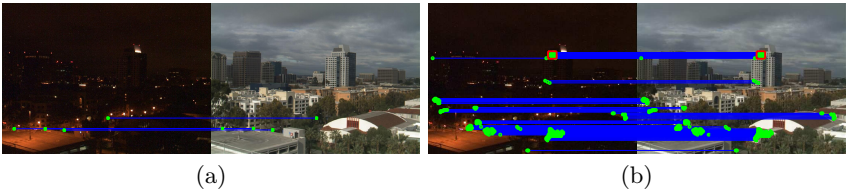

(a)   (b)

Fig. 9. (a) correct matches of DoG+RootSIFT. (b) correct matches of dense RootSIFT.

lies outside the region from which nearest neighbor's RootSIFT feature is extracted to avoid comparing similar features. Figure 8 (a) and (b) show the precision of dense RootSIFT matching and the number of matched feature points. Though the precision is not improved compared with the best performing detector+RootSIFT, the number of matched feature points improves a lot. This means that there are a lot of "easy" RootSIFT features that could be matched for day-night image pairs. However, we find that the matched RootSIFT features tend to cluster. Since detectors would usually perform non-maximum suppression to get stable detections, in the worst case, only one feature could be detected from each cluster. As a result, the number of these matched features is an upper bound that cannot be reached. Instead, we try to get a lower bound on the number of additional potential matches that could be found. To achieve that, we count the number of connected components for those matched RootSIFT features and show the result in Fig. 8 (c). Taking DoG as an example, we show the number of connected components that contain no correct matches found by detector+RootSIFT in Fig. 8 (d). We found that matches found by detector+RootSIFT have almost no overlap with the connected components of the matched dense RootSIFT, meaning that there is great potential to improve feature detectors. Moreover, we notice that there are generally 10 - 20 connected components found by dense RootSIFT. This is in the order of correct matches we could get for day-night image matching shown in Fig. 4.

Fig. 9 shows an example of correct matches found by DoG+RootSIFT and dense RootSIFT. For this day-night image pair, DoG+RootSIFT can only find 4
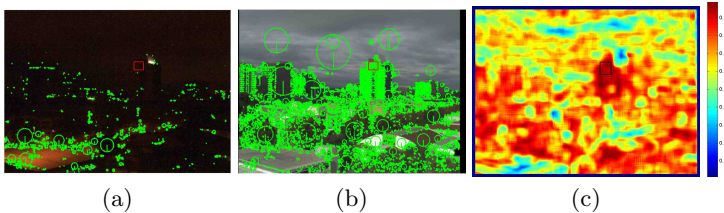
(a)   (b)   (c)

Fig. 10. (a) and (b) shows an example of nighttime and daytime image with detected feature points using DoG. (c) shows the heat map of the cosine distance of dense RootSIFT for (a) and (b).

correct matches whereas dense RootSIFT can find 188 correct matches. Fig. 10 shows the detected feature points using DoG for the day and night images and their corresponding heat map of cosine distance of dense RootSIFT. The colored rectangles in Fig. 9 (b) and those in Fig. 10 (a), (b) (c) are the same area. It is clearly shown that the cosine distances of points in that area between day and night images are very large, and Fig. 9 (b) shows that they can be matched using dense RootSIFT. However, though many feature points can be detected in the daytime image, no feature points are detected by DoG for the nighttime image[3]. As a result, matches that could be found by RootSIFT are missed due to the detector. In conclusion, a detector which is more robust to severe illumination changes can help improve the performance of matching day-night image pairs.

5   Conclusion

In this paper, we evaluated the performance of local features for day-night image matching. Extensive experiments show that repeatability alone is not enough for evaluating feature detectors. Instead, descriptors should also be considered. Through the discussion about precision and recall of matching day-night images and examining the performance of dense feature matching, we concluded that there is great potential for improving both feature detectors and descriptors. Thus, further evaluation with parameter tuning and advanced descriptors [11] as well as principal research on the day-night matching problem is needed.

References

1. Arandjelović, R., Zisserman, A.: Three things everyone should know to improve object retrieval. In: CVPR (2012)
2. Brown, M., Hua, G., Winder, S.: Discriminative learning of local image descriptors. IEEE Transactions on PAMI 33(1) (2011)
3. Furukawa, Y., Ponce, J.: Accurate, dense, and robust multiview stereopsis. IEEE Transactions on PAMI 32(8), 1362–1376 (2010)

---

[3] The area in the rectangle of the night image actually has a lot of structure, it appears to be totally dark due to low resolution

4. Han, X., Leung, T., Jia, Y., Sukthankar, R., Berg, A.C.: Matchnet: Unifying feature and metric learning for patch-based matching. In: CVPR (2015)
5. Jacobs, N., Roman, N., Pless, R.: Consistent temporal variations in many outdoor scenes. In: CVPR (2007)
6. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. IJCV 60(2) (2004)
7. Lynen, S., Sattler, T., Bosse, M., Hesch, J., Pollefeys, M., Siegwart, R.: Get out of my lab: Large-scale, real-time visual-inertial localization. In: RSS (2015)
8. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide baseline stereo from maximally stable extremal regions. In: BMVC (2002)
9. Mikolajczyk, K., Schmid, C.: An affine invariant interest point detector. In: ECCV (2002)
10. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Gool, L.V.: A comparison of affine region detectors. IJCV 65(1/2) (2005)
11. Mishkin, D., Matas, J., Perdoch, M., Lenc, K.: WxBS: Wide Baseline Stereo Generalizations. In: BMVC (2015)
12. Morel, J.M., Yu, G.: Asift: A new framework for fully affine invariant image comparison. SIAM J. Img. Sci. 2(2) (2009)
13. Richardson, A., Olson, E.: Learning convolutional filters for interest point detection. In: ICRA (2013)
14. Simo-Serra, E., Torras, C., Moreno-Noguer, F.: Dali: Deformation and light invariant descriptor. IJCV 115(2) (2015)
15. Simo-Serra, E., Trulls, E., Ferraz, L., Kokkinos, I., Fua, P., Moreno-Noguer, F.: Discriminative Learning of Deep Convolutional Feature Point Descriptors. In: ICCV (2015)
16. Simonyan, K., Vedaldi, A., Zisserman, A.: Learning local feature descriptors using convex optimisation. IEEE Transactions on PAMI 36(8) (2014)
17. Sivic, J., Zisserman, A.: Video Google: A Text Retrieval Approach to Object Matching in Videos. In: ICCV (2003)
18. Snavely, N., Seitz, S.M., Szeliski, R.: Photo tourism: exploring photo collections in 3d. In: ACM transactions on graphics (TOG). vol. 25, pp. 835–846 (2006)
19. Suenderhauf, N., Shirazi, S., Jacobson, A., Dayoub, F., Pepperell, E., Upcroft, B., Milford, M.: Place recognition with convnet landmarks: Viewpoint-robust, condition-robust, training-free. In: RSS (2015)
20. Torii, A., Arandjelović, R., Sivic, J., Okutomi, M., Pajdla, T.: 24/7 place recognition by view synthesis. In: CVPR (2015)
21. Triggs, B.: Detecting keypoints with stable position, orientation, and scale under illumination changes. In: ECCV (2004)
22. Trzcinski, T., Christoudias, M., Lepetit, V., Fua, P.: Learning image descriptors with the boosting-trick. In: NIPS (2012)
23. Vedaldi, A., Fulkerson, B.: VLFeat: An open and portable library of computer vision algorithms. http://www.vlfeat.org/ (2008)
24. Verdie, Y., Yi, K.M., Fua, P., Lepetit, V.: TILDE: A Temporally Invariant Learned DEtector. In: CVPR (2015)
25. Wu, C., Clipp, B., Li, X., Frahm, J.M., Pollefeys, M.: 3d model matching with viewpoint-invariant patches (vip). In: CVPR (2008)
26. Zagoruyko, S., Komodakis, N.: Learning to compare image patches via convolutional neural networks. In: CVPR (2015)
27. Zeisl, B., Sattler, T., Pollefeys, M.: Camera Pose Voting for Large-Scale Image-Based Localization. In: ICCV (2015)