



Evaluation of Feature Detectors, Descriptors and Match Filtering Approaches for Historic Repeat Photography

Ann-Katrin Becker^(✉) and Oliver Vornberger

Media Computer Science Group, Osnabrueck University,
Wachsbleiche 27, 49069 Osnabrueck, Germany
ann_katrin.becker@uos.de

Abstract. This work analyzes the suitability of classic feature detectors and descriptors as well as different match filters for matching historical to modern images. A variety of prominent detector and descriptor combinations are evaluated on a new dataset, that is composed of repeat photographs of various scenes exposed to tremendous change across years. Results show that a dense keypoint sampling is more effective than classic feature detection, while several descriptors achieve comparable performances. Yet, more important is an adequate match filtering approach that identifies correct correspondences despite their large descriptor distances. We show that the standard ratio test is unsuitable for matching historic to modern image pairs, while other filters based on keypoint geometry boost performance. Finally, we establish a complete pipeline including detectors, descriptors and match filtering methods. Our pipeline shows high performance on challenging image pairs beyond repeat photography, as our tests on another dataset show.

Keywords: Matching · Historic repeat photography · Benchmark

1 Introduction

Repeat photography, or rephotography, is the process of retaking an old photograph from the same viewpoint. The resulting picture pair provides a compelling visualization of the respective location's change throughout the years. In ecology repeat photography is used to document vegetation and climate change [9, 20], glacier melting [12] and geological erosion [13], while sociologists use it to study urban development and social change [31]. However, even for experienced rephotographers recapturing the original view “by eye” is often tedious and time consuming. In practice there are two approaches to support this process. The first is to match the original and its approximate rephotograph via postprocessing, while the second supports the process of recovering the original viewpoint

[5]. The full automation of both processes requires the solution to a single underlying challenge. This is the automatic identification of similarities between the original typically historic image and the modern scene. Due to dramatic variations in lightning and weather, high occlusion and different acquisition devices, this is a challenging problem. Examples are shown in Fig. 1.

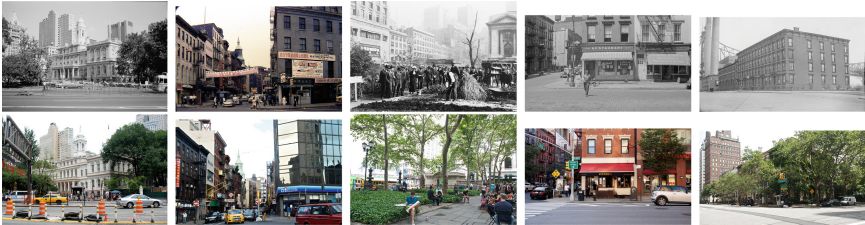


Fig. 1. Rephotographies from our dataset, including a view of a single building, a street corner lined with houses and a park. All image pairs feature high occlusion due to scene changes, show variations in lightning, and have been acquired by different cameras.

Feature matching is widely used for image alignment as well as 3D reconstruction. However, diverse opinions exist whether classic feature detectors such as SIFT [18] are suitable in the context of historical repeat photography. Bae et al. [5] and Schindler and Dellaert [25] agree that SIFT fails to match historic and modern images. Ali and Whitehead [2] instead, claim high performance on historical and modern image matching for at least some classic detector and descriptor combinations. Datasets from other studies do not adequately replicate the conditions present in historic repeat photography [14, 27]. Overall no detailed study exists that allows a conclusion whether classic feature detectors and descriptors are suitable for historic to modern image matching.

This paper presents such a detailed study. First of all we present a new dataset composed of publicly available rephotographs of Manhattan. This does not only contain images of single buildings, but also whole streets lined with houses that are subject to great change across the years. Secondly we analyze the suitability of a variety of feature detector and descriptor combinations. Afterwards the top performing detector and descriptor pairs are analyzed in more detail and their parameters are varied to optimize performance. Furthermore we apply different match filtering approaches to improve the low precision values faced with most detector and descriptor combinations.

From our detailed analysis, we draw the following conclusions: (1) We agree with Fernando et al. [10] and Stylianou et al. [27] and propose a dense sampling of keypoints instead of the application of classic detectors. (2) RootSIFT and SURF but also binary descriptors such as BRIEF [8] and LATCH [16] are suitable for matching historic and modern images. For the latter we recommend increasing descriptive power by extending descriptor length. (3) Due to high descriptor distances of correct matches, classic filtering approaches such as the ratio test

[18] fail in the context of historic repeat photography. Instead the application of other filtering approaches based on keypoint geometry boosts performance.

Finally we successfully apply our established feature detection, description and match filtering pipeline to the dataset of Hauagge and Snavely [14]. Thus the approach we present in this paper is applicable to challenging image pairs beyond rephotography, including day and night shots and images of varying rendering style.

2 Related Work

Few research concerned with matching historical to modern images has been presented in the literature. Closely related to this work is the evaluation of Gat et al. [11] who compared a set of classic feature detectors on repeat photographs of various mountain landscapes. Unfortunately, no different feature descriptors are assessed, though other works have shown that often specific detector and descriptor combinations outperform others [10, 14, 27]. Ali and Whitehead [2] on the other hand evaluate various detector and descriptor combinations on their own dataset, including images of popular sights across different time periods. Their results depict that even the most successful ORB/SURF combination often fails for matching historic and modern images separated by large time spans.

Bae et al. [5], who present a software prototype that supports recapturing of photograph, state that SIFT fails to identify similarities between historic and modern image pairs and allocate this task to the user. A similar statement is made by Schindler and Dellaert [25], who sort a collection of historic images with the help of manual user interaction.

Stylianou et al. [27] analyse feature matching performance over time periods of five years. They show that the influence of structural changes on matching is tiny compared to changes in weather and lightning. Yet, five years are a rather short time span in historic repeat photography. Similarly Valgren and Lilienthal [30] study the impact of seasonal changes upon localization in outdoor environments over a period of nine months. Other works [6, 14] propose entirely new features for challenging image pairs, including historic repeat photographs, as well as day and night images and images of different rendering styles.

In image retrieval further works on matching disparate images, including historic ones, exist. Shrivastava et al. [26] developed a method for cross domain image matching based on densely sampled HOG features with learned weights. Similarly Aubry et al. [4] use the HOG descriptor to align images from multiple domains to a 3D model. Fernando et al. [10] aim at recognizing the location of an old photograph given modern labeled images from the Internet. Torii et al. [28] address the challenge of large-scale place recognition on day and night images from modern day Tokyo, Both [10, 28] propose a combination of dense sampling and RootSIFT for feature detection and description. Anyhow, it remains unknown whether image representation and classifier training on large datasets compensate for poor actual descriptor performance.

3 Evaluation

In this paper we analyze the suitability of classic detectors and descriptors for matching historic and modern image pairs as they arise during historic repeat photography. A majority of previous studies revealed that such pairs pose a challenge to classic detectors and descriptors. Hence, our evaluation not only tests general performance but also analyses measures to improve it. These include varying detector and descriptor parameters as well as applying different match filtering techniques. To compare these measures for a large set of detector and descriptor combinations we use a two step approach. At first we perform an initial evaluation applying a large amount of detector and descriptor combinations to our new dataset using default parameters. Afterwards a detailed evaluation follows, where we improve matching performance for the top detector and descriptor pairs. In the following we provide an overview of our dataset, the evaluation criteria, tested detectors and descriptors and match filters.

3.1 Dataset

Since the datasets used in related studies do not meet the conditions of this work [14], are not publicly available [11], or both [2], we constructed our own dataset for this study, which is available online¹. This contains 52 rephotographies of Manhattan from a collection created by Paul Sahner². The modern photographs were taken between 2009 and 2013, while the originals span the whole 20th century. The collection contains a variety of scenes including single buildings, as well as views of park areas or whole streets lined with houses. For some examples refer to Fig. 1. Furthermore the scenes are exposed to different amounts of structural change across the years. Additional variations between image pairs are caused by changes in lightning and weather as well as different acquisition devices or rendering styles.

As taken by a professional rephotographer most of the image pairs of the collection are already well aligned. Yet for further improvement we additionally computed a homography between all image pairs based on manual selection.

3.2 Evaluation Criteria

Related studies use *repeatability*, *precision*, *recall* [14] and *pass rate* [11] to evaluate detector and descriptor performance.

Repeatability assesses feature detector performance. For each detected keypoint the ground truth homography is used to measure whether a corresponding keypoint at the same location and of similar size exists [21]. In the context of rephotography often structures disappear and thus their features can no longer be repeated in the modern image. Instead accidental correspondences arise, since new structures feature new keypoints. Thus standard repeatability measures

¹ <https://github.com/AnnKatrinBecker/nyc-grid-dataset>.

² <http://www.paulsahner.com>, <http://www.nyc-grid.com>.

many more keypoint correspondences than can actually be returned during the descriptor based matching process. So that reachable recall rates (number of correct matches divided by the number of repeated keypoints) drop far below 100%. To establish a reliable ground truth for repeatability in the context of historic and modern images requires to annotate each region suffering from occlusion manually. Thus we refrain from using repeatability and recall in our evaluation.

Precision depicts the ratio of inliers among all corresponding keypoints returned by the matching process. Compared to repeatability this inlier rate suffers from very few false positives since it is reliant on descriptor distance. *Average precision* depicts the average precision value among the whole dataset. Unfortunately this value does not provide any information on the precision distribution.

Pass rate [11] instead, is calculated on the entire dataset and requires minimum thresholds for precision and the number of correct matches. Then pass rate is the percentage of image pairs from the dataset that exceed the appointed thresholds. Especially if few images show a high number of correct correspondences, while others hardly contain any, pass rate is more meaningful than average precision. Thus we prefer pass rate in our evaluation.

3.3 Detectors and Descriptors

We aimed at including all detectors and descriptors in our test set that showed good performance in related studies, including dense sampling [10, 14, 28], Root-SIFT [10, 28] and the combinations of SURF/U-SURF [30], SIFT/SIFT [14] and ORB/SURF [2]. Afterwards we extended it with prominent detectors such as MSER and FAST as well as features only recently proposed (e.g. LATCH, AKAZE). Furthermore our test set includes fast binary as well as more computational expensive floating point descriptors.

In summary the detectors compared in this paper include dense sampling, MSER [19] and FAST [22]. Tested descriptors are U-SURF [7], RootSIFT [3], BRIEF [8] and LATCH [16]. Moreover the following features are used as detectors and descriptors: SIFT [18], SURF [7], ORB [24], BRISK [15], AKAZE [1]. We use the OpenCV implementation of all these features, including their respective default parameters. Only for ORB we limit the maximum number of keypoints to 10000 instead of 500.

3.4 Match Filtering Approaches

For match filtering we use classic approaches such as the ratio test [18] and maximal descriptor distance. However, as correct matches feature large descriptor distances in the context of historic and modern image pairs, descriptor distance is not a well distinguishing feature. Thus we additionally evaluate filtering approaches beyond descriptor distance. This includes the disparity gradient filter (DGF) proposed by Roth and Whitehead [23], exclusively based on the geometry of keypoints, as well as K-VLD [17], that evaluates both geometric as well as photometric consistency in the local neighborhood of a keypoint pair.

Disparity Gradient Filter. Given an image pair the disparity gradient measures the geometric compatibility of two of its correspondences [29]. Let P_1 and P_2 be keypoints in the first image and $f(P_1)$ and $f(P_2)$ be their correspondences in the second image, then their disparity gradient

$$d = \frac{|(P_1 - P_2) - [f(P_1) - f(P_2)]|}{\frac{1}{2}|(P_1 - P_2) + [f(P_1) - f(P_2)]|} \tag{1}$$

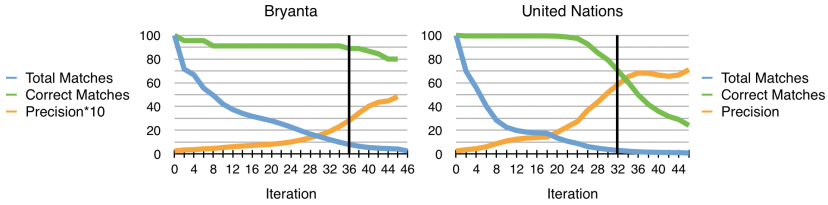


Fig. 2. Detailed course of applying the disparity gradient filter to individual image pairs. Both graphs illustrate the severe drop in total matches, while only few correct matches are discarded. This boosts precision. Beyond our point of termination (vertical line), precision values increase, but the loss of correct matches becomes more significant.

In case two correspondences conform well their disparity gradient is small. For general image registration Roth and Whitehead [23] sum up the disparity gradient for each correspondence with respect to each other correspondence. Afterwards correspondences with a high *DGS* are removed. Both steps are performed iteratively until the *max DGS* is less than twice the *min DGS*.

In general we follow the approach of Roth and Whitehead [23], while we apply a softer termination criterion and discard false matches by the following rule. During each iteration we determine the *median DGS* and discard all correspondences featuring a *DGS* greater than the median times a certain factor. Initially this factor is set to $1\frac{1}{2}$. Afterwards it decreases to $1\frac{1}{4}$, $1\frac{1}{8}$ and so on if

$$factor * median \geq max DGS \tag{2}$$

Furthermore we decrease the factor if less than 1% of correspondences were eliminated during the last iteration. During initial testing we also realized that the termination criterion of $min DGS * 2 > max DGS$ is to tight for our image pairs. Thus after analyzing initial results, see Fig. 2 we changed it to

$$min DGS * 3 > max DGS \tag{3}$$

4 Results

4.1 Initial Evaluation

In the initial evaluation we assess the performance of diverse feature detectors and descriptors on historic and modern image pairs. We apply all possible combinations of the feature detectors and descriptors listed in Sect. 3.3 to our new

dataset. In the case of dense sampling, we apply two variants named Dense10 (dense sampling with a step size and a keypoint radius of 10) and Dense25 (step size and radius equal 25 as proposed in [14]). For matching keypoints nearest neighbor search is performed using the hamming distance or L2 norm, depending on the descriptor used.

Table 1. Set of top ten detector and descriptor combinations after the initial evaluation. All combinations are ranked by pass rate 16 (last column).

Detector	Descriptor	Prec.(%)	pr16(%)	Detector	Descriptor	Prec.(%)	pr16(%)
ORB	RootSIFT	3.9	96	Dense25	RootSIFT	12.0	83
FAST	RootSIFT	3.3	92	Dense10	LATCH	2.7	83
SURF	RootSIFT	3.9	90	ORB	BRIEF	2.4	83
Dense10	RootSIFT	3.8	88	FAST	BRIEF	1.7	83
ORB	U-SURF	3.1	88	FAST	LATCH	2.5	81

Previous studies report that it is critical to find enough correct matches at all for rephotographic image pairs [25]. Thus, initially we refrain from applying any filters, since these also eliminate correct matches. Of course this leads to poor precision values. Consequently in the initial evaluation we only use the number of correct matches to compare feature detectors and descriptors with each other, while we expect to improve precision during the detailed evaluation.

Therefore, we compute pass rates requiring a minimum number of 16 correct correspondences, while imposing no threshold on precision. We choose 16 as minimum value, since it is the required minimum number of correct correspondences in [25]. Table 1 depicts the top ten detector and descriptor combinations sorted by pass rate 16. Note that RootSIFT showed higher or similar performance as SIFT for all relevant combinations, thus we omit displaying the results of SIFT. Our further detailed evaluation tried to optimize the performance of these top combinations.

4.2 Varying Detector and Descriptor Parameters

At first we vary the parameters of dense sampling. Initially we keep the spacing of 10 pixels and sample dense keypoints with radii of 25, 35 and 50 pixels. While increasing the keypoint radius to 25 has a positive effect on pass rate, further increasing the radius to 35 or 50 does not improve but rather diminishes pass rate. Secondly we decrease keypoint spacing to only 5 pixels so that the total number of keypoints sampled is increased. Especially in combination with LATCH this shows a positive impact on pass rate as well, while there is no significant negative effect on precision.

Thus we decided to consider DENSE5:25/RootSIFT and DENSE5:25/LATCH in our further evaluation. However, we note that dense sampling requires

image pairs of similar scale and ideal keypoint scale is related to image resolution. Hence, if image resolution varies from our dataset (around 800×600 pixel) other radii may show better performances.

In the case of dense sampling raising the number of keypoints increased performance. Thus, we also analyze the effects of varying keypoint number for other detectors among the top ten. For ORB we evaluate a restriction of keypoints to 10000, 15000 and 20000, while for SURF we vary the Hessian Threshold from 100 to 50 and 25. In general an increase in the number of sampled keypoints results in an increase in pass rate, while precision values are hardly effected. Only for the combination ORB/BRIEF sampling 20000 keypoints diminishes performance. We attribute this effect towards the lower descriptive power of BRIEF as a binary descriptor compared to its floating point counter parts RootSIFT and U-SURF. Furthermore in the case of SURF detector a Hessian threshold of 50 depicts the best performance. Thus we use ORB20k/RootSIFT, ORB20k/U-SURF, ORB15k/BRIEF and SURF/RootSIFT with a hessian threshold of 50 for our further evaluation. Since FAST already samples around 15000 keypoints on each image, we did not change its parameters further.

At last we vary descriptor length to enhance discriminative power, especially for binary descriptors. For BRIEF and LATCH we compare descriptor lengths of 16, 32 and 64 bytes, while for U-SURF we compare a descriptor lengths of 64 and 128 floating point numbers. As expected for binary descriptors increasing descriptor length shows positive effects on pass rate and precision. For U-SURF instead changing descriptor length has hardly any effect.

In summary, we remain with the following nine detector and descriptor combinations for our further evaluation: Dense5:25/RootSIFT, Dense5:25/LATCH64, ORB20k/RootSIFT, ORB20k/U-SURF, FAST/RootSIFT, SURF/RootSIFT, FAST/BRIEF64, ORB15k/BRIEF64, FAST/LATCH64.

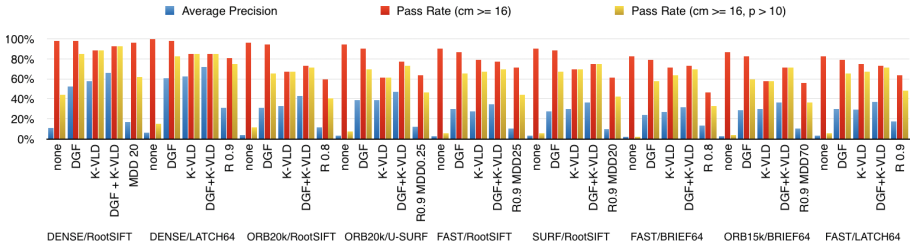


Fig. 3. Illustration of the performance of different match filtering approaches. In detail for each detector and descriptor combination results of applying no filter, DGF, K-VLD, the combination of both and the top performing standard filter (ratio test, maximal descriptor distance, or their combination) are displayed. For most detector and descriptor combinations pure K-VLD performs a little poorer than DGF in terms of pass rate, while similar average precision values are reached. Overall applying DGF followed by K-VLD, shows the best performance for most combinations.

4.3 Match Filtering

At first we applied the standard ratio test, with thresholds of 0.9, 0.8 [18] and 0.6 [25] for match filtering. However, only a weak threshold of 0.9 leads to slight performance increase in terms of precision and pass rate. For stricter values there still is an increase in precision, but this is accompanied by a reduction of correct correspondences below 16 for a significant number of image pairs.

Secondly, we evaluated the effects of applying a maximal descriptor distance threshold. Results are similar to those of applying the ratio test. Only very soft thresholds lead to slight increases in pass rate, while tighter ones diminish too many correct matches. Also a combination of both, ratio test and maximal descriptor distance threshold, does not result in promising increases in pass rate if a precision above 10% is required. Exemplary results are illustrated in the last column of Fig. 3.

Thirdly we applied the disparity gradient filter (DGF) [23] with our previously described iteration method and termination criterion. Our evaluation shows that DGF hardly discards any correct matches while lots of false correspondences are eliminated. Hence, DGF is able to boost precision by factors of 10 to 30, even for extremely low initial precision values beyond 1%, see also Fig. 2. The number of correct matches on the other hand does not diminish significantly upon DGF application.

At last we evaluated the performance of K-VLD [17] as a match filter for historic and modern image pairs. Overall K-VLD performs a little poorer than DGF in terms of pass rate, even though average precision is similar or slightly higher than for DGF. For a direct comparison please refer to Fig. 3. For individual images K-VLD leads to a greater boost in precision, but it also discards more correct matches. Hence for some image pairs its application results in a drop of correct matches below 16.

Due to the individual shortcomings of DGF (low precision) and K-VLD (few correct matches) for individual image pairs, we propose to combine both filters. At first DGF is used to discard a majority of false correspondences and reach precision values above 1% or 2%. Afterwards K-VLD is applied to achieve a further boost in performance. The results of applying this filter combination to our dataset are displayed in the fourth column of Fig. 3. Compared to single DGF application a clear performance increase in average precision is visible, while such is not accompanied by a drop in pass rate as in the case of single K-VLD use.

In summary the generally low performance of standard filters, such as the ratio test and maximal descriptor distance, can be ascribed to their dependence on descriptor distance. This is, since for image pairs featuring major appearance changes, even correct matches feature high descriptor distances. Thus descriptor distance is not a distinctive feature between correct and false correspondences of modern and historic image pairs. Instead, measures such as DGF, exclusively based on keypoint geometry, and K-VLD, based on photometric as well as geometric measures, are much more effective for filtering correspondences.

4.4 Application Beyond Rephotography

Finally we verify the applicability of our established pipeline to other challenging image pairs beyond rephotography and our own dataset. To do so we apply our top nine detector and descriptor combinations together with our proposed filtering mechanisms to the dataset of Hauagge and Snavely [14]. This is composed of various challenging image pairs, including day and night images and images of different rendering styles.

Table 2. Comparison with the results of Hauagge and Snavely [14]

(a) Mean average precision of [14].

	GRID	SIFT	SYM-I	SYM-G
Self-Sim.	0.29	0.14	0.12	0.16
SIFT	0.49	0.21	0.28	0.25
SYMD	0.41	0.22	0.20	0.25
SIFT-SYMD	0.58	0.28	0.35	0.36

(b) Average precision of our pipeline.

	none	DGF	K-VLD	both
Dense5:25/RootSIFT	0.24	0.69	0.79	0.84
Dense5:25/LATCH64	0.13	0.71	0.77	0.85
ORB20k/RootSIFT	0.11	0.56	0.58	0.71
ORB20k/U-SURF	0.10	0.64	0.65	0.74

Table 2 depicts the average precision of our top four combinations, without match filtering, after applying DGF or K-VLD and the combination of both. Overall the average precision values reached on the dataset of Hauagge and Snavely [14] are higher than those achieved for our own dataset reported in Fig. 3. Hence as expected, our dataset is more challenging due to the presence of major structural changes.

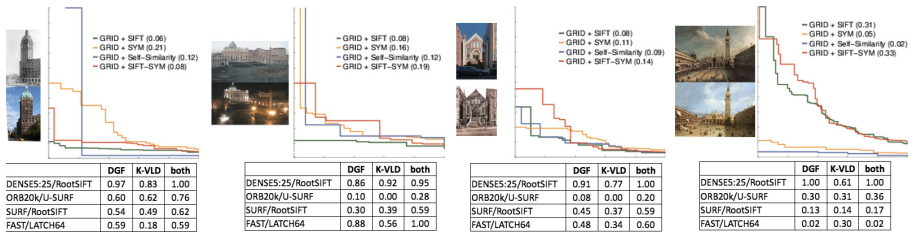


Fig. 4. Comparison of results for individual image pairs from [14]. At the top the image pairs and their respective recall and precision curves for dense sampling as reported by [14] are displayed. At the bottom the precision curves values of a variety of our detector and descriptor combination reached after match filtering are shown. For the majority of image pairs our proposed pipeline reaches high precision values

The mean average precision values reached by [14] are repeated in Table 2. Unfortunately, their mean average precision values are not directly comparable to our average precision reached after filtering, since they applied a variety of ratio thresholds for filtering and report mean average precision of all these tests.

Anyhow, there is a distance of more than 0.25 between their mean average precision values for dense sampling (GRID) and the average precision values after DGF and K-VLD application for our top two combinations using dense sampling. This suggests that our proposed pipeline outperforms the approach of [14] in matching image pairs containing high appearance variations. To confirm this we also compared the performance of both approaches on individual image pairs, which have shown to be extremely challenging based on the results of [14]. A subset of these image pairs is displayed in Fig. 4. These examples show that for a majority of image pairs, that feature rather low precision and recall curves in [14], our pipeline is able to generate high precision values after match filter application.

5 Conclusion

In this work we analyzed the suitability of classic feature detectors and descriptors for matching historical and modern image pairs. In this context we created a new challenging publicly available dataset, composed of rephotographies of a variety of scenes exposed to moderate up to tremendous change across years. Our evaluation results show that selective feature detection is not suitable if image pairs contain great variations. More effective is a dense sampling of keypoints. As descriptors overall RootSIFT and SURF but also the binary descriptors BRIEF and LATCH show the highest feature matching performance. Furthermore we discovered that a suitable match filtering approach is much more important than the individual classic detector and descriptor combination used. In detail filters based on descriptor distance are not suitable in the presence of great appearance changes. Instead the application of filters based on geometry such as the disparity gradient filter [23] or K-VLD [17] boost performance.

In summary, we established a complete pipeline including detectors, descriptors and matching filtering methods. This also shows high performance in matching challenging image pairs beyond historic repeat photography as our comparison on another dataset showed. However, even after application of our pipeline, we only reach pass rates of 70% to 90% if we presume a minimum number of 16 correct matches and precision values above 10%, remember Fig. 3. Given these requirements, between 10% to 30% of image pairs of our dataset remain unmatchable. Consequently further research and possibly completely new feature matching approaches are required to match even the most challenging historic repeat photographs. Further more the performance of feature detectors and descriptors in more natural non urban environments still needs to be assessed.

References

1. Alcantarilla, P.F., Solutions, T.: Fast explicit diffusion for accelerated features in nonlinear scale spaces. *TPAMI* **34**(7), 1281–1298 (2011)
2. Ali, H.K., Whitehead, A.: Feature matching for aligning historical and modern images. *Int. J. Comput. Appl.* **21**(3), 188–201 (2014)

3. Arandjelović, R., Zisserman, A.: Three things everyone should know to improve object retrieval. In: CVPR, pp. 2911–2918 (2012)
4. Aubry, M., Russell, B.C., Sivic, J.: Painting-to-3D model alignment via discriminative visual elements. TOG **33**(2), 14 (2014)
5. Bae, S., Agarwala, A., Durand, F.: Computational rephotography. ACM Trans. Graph. **29**(3), 1–15 (2010)
6. Bansal, M., Daniilidis, K.: Joint spectral correspondence for disparate image matching. In: CVPR, pp. 2802–2809 (2013)
7. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006). https://doi.org/10.1007/11744023_32
8. Calonder, M., Lepetit, V., Strecha, C., Fua, P.: BRIEF: binary robust independent elementary features. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6314, pp. 778–792. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15561-1_56
9. Clements, F.E.: Research methods in ecology. The University Publishing Company, Lincoln (1905)
10. Fernando, B., Tommasi, T., Tuytelaars, T.: Location recognition over large time lags. Comput. Vis. Image Underst. **139**, 21–28 (2015)
11. Gat, C., Albu, A.B., German, D., Higgs, E.: A comparative evaluation of feature detectors on historic repeat photography. In: Bebis, G., et al. (eds.) ISVC 2011. LNCS, vol. 6939, pp. 701–714. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-24031-7_70
12. Gore, A.: An Inconvenient Truth: The Planetary Emergency of Global Warming and What We Can Do About It. Rodale, Emmaus (2006)
13. Hall, F.C.: Photo point monitoring handbook: part a-field procedures. Technical report, PNW: USDA Forest Service (2002)
14. Hauage, D.C., Snavely, N.: Image matching using local symmetry features. In: CVPR, pp. 206–213 (2012)
15. Leutenegger, S., Chli, M., Siegwart, R.Y.: Brisk: binary robust invariant scalable keypoints. In: ICCV, pp. 2548–2555 (2011)
16. Levi, G., Hassner, T.: LATCH: learned arrangements of three patch codes. In: WACV, pp. 1–9 (2016)
17. Liu, Z., Marlet, R.: Virtual line descriptor and semi-local matching method for reliable feature correspondence. In: BMVC, p. 16-1 (2012)
18. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
19. Mata, J., Chum, O., Urban, M., Pajdla, T.: Robust wide-baseline stereo from maximally stable extremal regions. Image Vis. Comput. **22**(10), 761–767 (2004)
20. Meyer, J.L., Youngs, Y.: Historical landscape change in yellowstone national park. Geogr. Rev. **108**(3), 387–409 (2018)
21. Mikolajczyk, K., et al.: A comparison of affine region detectors. Int. J. Comput. Vis. **65**(1–2), 43–72 (2005)
22. Rosten, E., Drummond, T.: Machine learning for high-speed corner detection. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 430–443. Springer, Heidelberg (2006). https://doi.org/10.1007/11744023_34
23. Roth, G., Whitehead, A.: Using projective vision to find camera positions in an image sequence. In: Conference on Vision Interface 2000, Canada, pp. 87–94 (2000)
24. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: An efficient alternative to SIFT and SURF. In: ICCV, pp. 2564–2571 (2011)

25. Schindler, G., Dellaert, F.: 4D cities: analyzing, visualizing, and interacting with historical urban photo collections. *J. Multimedia* **7**(2), 124–131 (2012)
26. Shrivastava, A., Malisiewicz, T., Gupta, A., Efros, A.A.: Data-driven visual similarity for cross-domain image matching. *TOG* **30**(6), 154 (2011)
27. Stylianou, A., Abrams, A., Pless, R.: Characterizing feature matching performance over long time periods. In: *WACV*, pp. 892–898 (2015)
28. Torii, A., Arandjelović, R., Sivic, J., Okutomi, M., Pajdla, T.: 24/7 place recognition by view synthesis. In: *CVPR*, pp. 1808–1817 (2015)
29. Trivedi, H.P., Lloyd, S.A.: The role of disparity gradient in stereo vision. *Perception* **14**(6), 685–690 (1985)
30. Valgren, C., Lilienthal, A.J.: SIFT, SURF & seasons: appearance-based long-term localization in outdoor environments. *Robot. Auton. Syst.* **58**(2), 149–156 (2010)
31. Walker, J., Leib, J.: Revisiting the Topia road: walking in the footsteps of West and Parsons. *Geogr. Rev.* **92**(4), 555–581 (2002)