ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Computer Vision and
Geometry Lab

Bachelor Thesis

A Radial Distortion Invariant Features Detector and Descriptor

Spring Term 2021

Supervised by:
Rémi Pautrat
Dr. Viktor Larsson
Prof. Dr. Marc Pollefeys

Author:
Sepideh Mamooler

Declaration of Originality

I hereby declare that the written work I have submitted entitled

A Radial DistortionInvariant Features Detectorand Descriptor

is original work which I alone have authored and which is written in my own words.[1]

Author(s)

Sepideh	Mamooler

co-supervisor(s)

Rémi	Pautrat
Viktor	Larsson

Supervising lecturer

Marc	Pollefeys

With the signature I declare that I have been informed regarding normal academic citation rules and that I have read and understood the information on 'Citation etiquette' (`https://www.ethz.ch/content/dam/ethz/main/education/rechtliches-abschluesse/leistungskontrollen/plagiarism-citationetiquette.pdf`). The citation conventions usual to the discipline in question here have been respected.

The above written work may be tested electronically for plagiarism.

Zurich, 31.08.2020	Sepideh Mamooler
Place and date	Signature

[1] Co-authored work: The signatures of all authors are required. Each signature attests to the originality of the entire piece of written work in its final form.

Contents

Abstract

Detecting interest points of images is the first step in geometric computer vision tasks such as Simultaneous Localization and Mapping (SLAM), Structure-from-Motion (SfM), camera calibration, and image matching. Interest points are 2D locations in an image which are stable and repeatable from different lighting conditions and viewpoints. The inputs to most real-world computer vision systems are raw images, not idealized point locations, and due to the relative motion between camera and the scene, these raw images have significant changes in feature appearance comparing to the original scene. There exists different algorithms that are invariant to affine transformations like scale and rotation and can partially overcome this problem. However, these approaches are not invariant to the radial distortion, or RD, in the images caused by cameras with wide field-of-view. This paper presents a modification to SuperPoint [1], a self-supervised framework for training interest point detectors and descriptors, that improves the repeatability of the detector and provides a more discriminating descriptor in radially distorted images. This improvement is obtained by distorting synthetic shapes in the pre-training phase and combining SuperPoint's homographic adaptation with RD-adaptation and generating the pseudo-ground truth for distorted images. Our model, when trained on the MS-COCO generic image dataset, results in a higher repeatability rate compared to the original Superpoint model and classical detectors like FAST, Harris, and Shi in the presence of radial distortion.

Related Work

Feature detection and description is a well studies area in the field of computer vision. The FAST corner detector [2] was the first system to cast high-speed corner detection as a machine learning problem. This method exploits the property of corners that the change of image instensity should be high in all direction. The Scale-Invariant Feature Transform, or SIFT [3], is still probably the most well-known traditional local feature descriptor in computer vision as it is invariant to ranslation, rotation, scale and low illumination changes. The Binary Robust Independent Elementary Features or BRIEF [4] is another feature descriptor that uses binary strings as an efficient feature descriptor and can be combine with arbitrary detectors. SURF [5] is also an interest point detection-description scheme which is fast and performant. Convolutional neural networks have been shown to be superior to hand-engineered representations on almost all tasks requiring images as input and interest point detection and description is not an exception to this fact. The Learned Invariant Feature Transform or LIFT [6], has a Deep Network architecture with a full feature point handling pipeline that learns to do interest point detection, orientation estimation and descriptor computation in a unified manner while preserving end-to-end differentiability. SuperPoint [1] is another example of these detectors. It is a fully-convolutional neural network architecture for interest point detection and description trained using a self-supervised domain adaptation framework called Homographic Adaptation. Since the release of SuperPoint the field of learned detectors and descriptors has exploded. D2-Net [7], R2D2 [8], and ASLFeat [9] are some examples of such methods. These methods are invariant to many image transformations like scale and rotation, but not to radial distortion.
Several methods including [10], [11], and [12] suggest correcting the radial distortion caused by camera lenses. Once the distortion is corrected any interest point detector and descriptor that is not RD-Invariant can be used for further image processing. However, this method adds another step to the pipeline, and having an interest point detector and descriptor that is invariant to radial distortion can be a better option. In this regard, sRD-SIFT [13] proposes modifications to SIFT [3], that improve the repeatability of detection and effectiveness of matching in the presence of distortion. These modifications require an approximate modeling of the image distortion, and consists in using adaptative gaussian filtering for detection and implicit gradient correction for description. There is also MDBRIEF [14] which is a fast online-adaptable, distorted binary descriptor for real-time applications using calibrated wide-angle or fisheye cameras. It proposes a distorted and masked version of the BRIEF descriptor [4] for calibrated cameras. Instead of correcting the distortion holistically, it distorts the binary tests and thus adapt the descriptor to different image regions.

Introduction

Interest point detection and description is one of the primary steps in all computer vision tasks. There exists several methods used for detecting interest points and computing their corresponding descriptors and SuperPoint is one of them.
SuperPoint is a self-supervised feature detector and descriptor that is invariant to common image transformations like scale, rotation and illumination. This mechansime consists of a pre-training phase in which it uses a set of labeled simple geometric shapes called synthtic dataset and warps them using random homographies to train a base detector. By using this base detector along with its homographic adaptation technique, SuperPoint self-labels the unlabeled input images and generates the pseudo-ground truth interest point for each image. Finally, this pseudo-ground truth is used to train the detector and descriptor. SuperPoint can detect interest points in real time with high repeatability and compute their corresponding descriptors with a high homography estimation rate. However, like many other state-of-the-art interest point detectors and descriptors, SuperProint is not robust to radial distortion present in the image caused by bending of the light rays when crossing the optics. Radial distortion is a non-linear deformation that moves the pixels of the image along the radiuses starting from the center of distortion. Most commercially available cameras introduce radial distortion in the images which leads to inaccurate interest point detection in many cases. Currently there exists methods that solve this problem by correcting the radial distortion present in the image and processing the corrected image based on the assumption of a linear pinhole camera. This requires every image to be corrected before being processed any further for interest point detection. As a result, detecting the interest points and computing their descriptors would no longer be possible in real time even when using methods like SuperPoint.
In this report, we show that by applying some modifications to the SuperPoint algorithm we can make it more robust to radial distortion and thus provide real time interest point detection and descriptor computation even in the presence of radial distortion. We introduce the RD-adaptation technique which is inspired by the homographic adaptation used in SuperPoint, and show that by combining these two techniques we can achieve more robustness in detecting interest points of distorted images. In the first chapter we explain SuperPoint's mechanism. The second chapter describes how we can distort images before using them in the pre-training phase and use RD-adaptation during the self-labeling process of SuperPoint. Finally, we show the results of evaluating our RD-invariant feature detector and descriptor and compare it with existing algorithms both for distorted images and images with no distortion.

Chapter 1

SuperPoint

1.1 SuperPoint Architecture

SuperPoint [1] is a self-supervised framework for training interest point detectors and descriptors. It is a fully-convolutional model which operates on full-sized images and jointly computes pixel-level interest point locations and associated descriptors in a single forward pass (see Figure 1.1). SuperPoint is suitable for a large number of multiple-view geometry problems in computer vision, and outperforms classical feature detectors and descriptors like SIFT, LIFT, and ORB in standard evaluation metrices for feature detectors and descriptors including repeatability and homography estimation.

On the contrary to the other feature detectors and descriptors, SuperPoint tackles interest point detecting and computing their corresponding descriptors in a single network. The detector and descriptor networks share a VGG-style [15] encoder to process and reduce the input image dimensionality. Then the network splits in two decoder heads which learn task specific weights – one for interest point detection and the other for interest point description. For a more detailed explanation of the decoder heads please refer to [1].

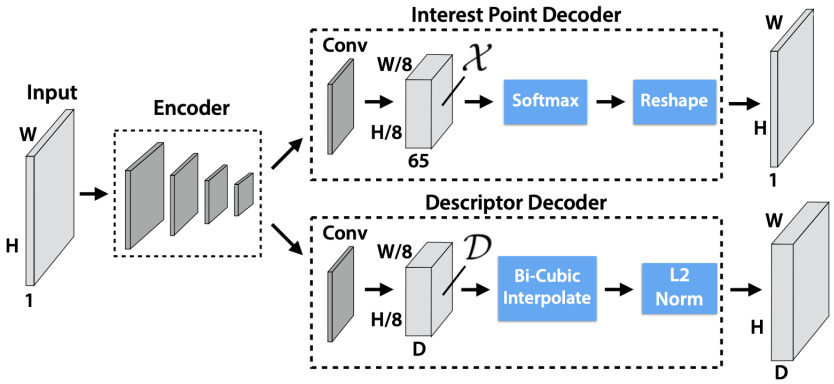

Figure 1.1: **SuperPoint Decoders.** Both decoders operate on a shared and spatially reduced representation of the input. To keep the model fast and easy to train, both decoders use non-learned upsampling to bring the representation back to $\mathbb{R}^{H \times W}$. (taken from [1]).

1.2 MagicPoint

Instead of using human supervision to define interest points in real images, SuperPoint presents a self-supervised solutiong. It creates a large dataset of pseudo-ground truth interest point locations in real images, supervised by the interest point detector itself, rather than a large-scale human annotation effort. Figure 1.2 shows an overview of this self-supervised training.

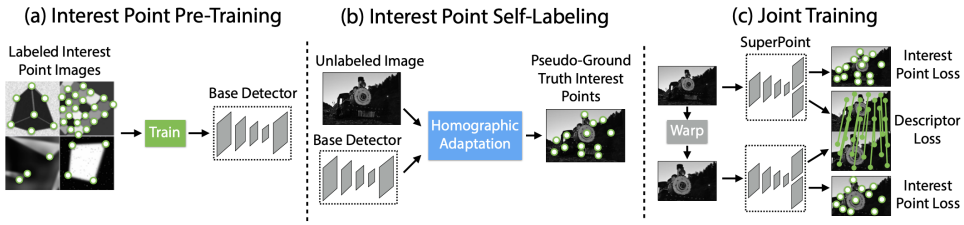

Figure 1.2: **Self-Supervised Training Overview.** In its self-supervised approach, SuperPoint (a) pre-trains an initial interest point detector on synthetic data and (b) applies a novel Homographic Adaptation procedure to automatically label images from a target, unlabeled domain. The generated labels are used to (c) train a fully-convolutional network that jointly extracts interest points and descriptors from an image. (taken from [1])

To generate the pseudo-ground truth interest points, a fully-convolutional neural network is trained on millions of examples from a synthetic dataset called Synthetic Shapes which are simple geometric shapes with no ambiguity in the interest point locations. The resulting trained detector is called MagicPoint. It significantly outperforms traditional interest point detectors on the synthetic dataset. (See Figure 1.3).

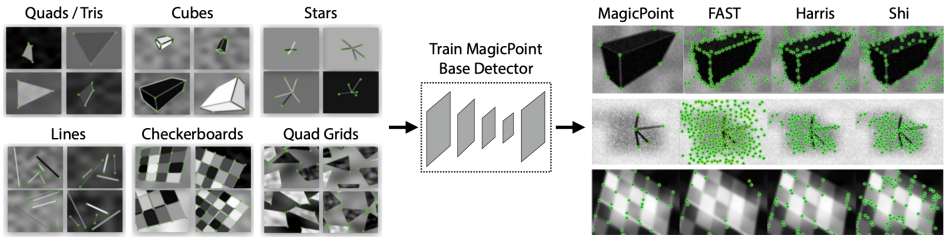

Figure 1.3: **Synthetic Pre-Training.** SuperPoint uses Synthetic Shapes dataset consisting of rendered triangles, quadrilaterals, lines, cubes, checkerboards, and stars each with ground truth corner locations. The dataset is used to train the MagicPoint convolutional neural network, which is more robust to noise when compared to classical detectors. (taken from [1])

The input to the computer vision tasks are real world images with various textures and patterns rather than simple geometric shapes present in the synthetic dataset. Although MagicPoint performs surprising well on real images despite domain adaptation difficulties [16], when compared to classical interest point detectors on a diverse set of image textures and patterns, it misses many potential interest point locations. SuperPoint's solution to this gap in performance on real images, is a technique called Homographic Adaptation.

1.2.1 Synthetic Shapes

There is no large database of interest point labeled images that exists today. Thus to bootstrap its deep interest point detector, SuperPoint first creates a large-scale synthetic dataset called Synthetic Shapes that consists of simplified 2D geometry via synthetic data rendering of quadrilaterals, triangles, lines and ellipses. Examples of these shapes as well as their interest points are shown in Figure 1.4. In this dataset, label ambiguity is removed by modeling interest points with simple Y-junctions, L-junctions, T-junctions as well as end points of line segments.

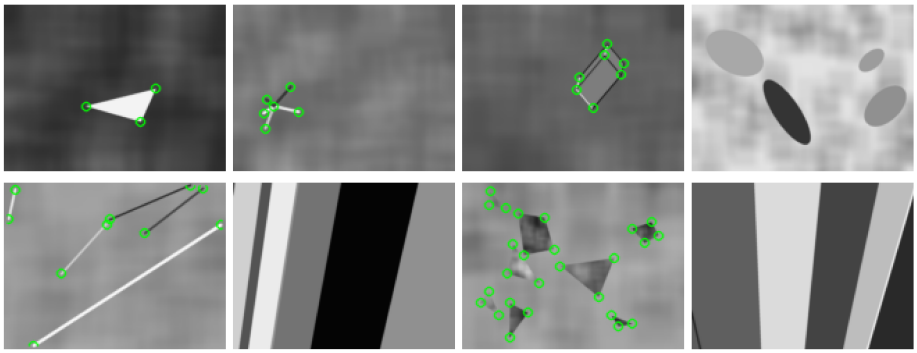

Figure 1.4: **Synthetic Shapes**. Synthetic shapes are simple geometric shapes with no ambiguity in the interest point locations. (Taken from SuperPoint's open source implementation of Paul-Édouard Sarlin and Rémi Pautrat [2].)

Once the synthetic images are rendered, homographic warps are applied to each image to augment the number of training examples. In Figure 1.5 some synthetic shapes warped with random homographies are shown with their keypoints. The data is generated on-the-fly and no example is seen by the network twice. While the types of interest points represented in Synthetic Shapes represents only a subset of all potential interest points found in the real world, it works reasonably well in practice when used to train an interest point detector.

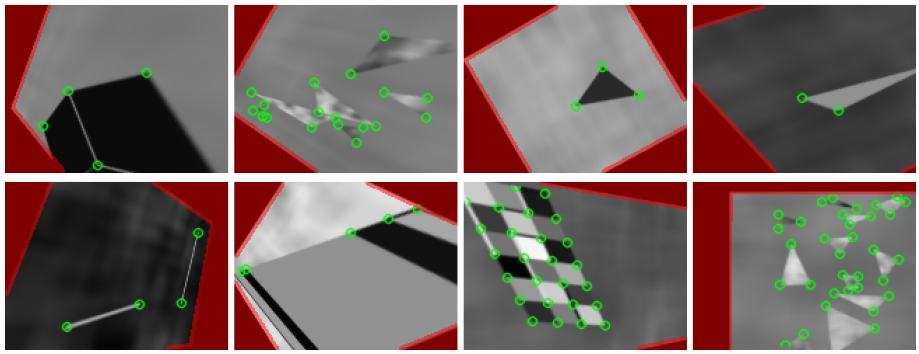

Figure 1.5: **Synthetic Shapes Warped with Random Homographies**. The red mask is used to indicate the image data lost after applying the homography.

[2] `https://github.com/rpautrat/SuperPoint`

1.2.2 Homographic Adaptation

Homographic Adaptation is designed to enable self-supervised training of interest point detectors. It warps the input image multiple times to help an interest point detector see the scene from many different viewpoints and scales. SuperPoint uses Homographic Adaptation in conjunction with the MagicPoint detector to boost the performance of the detector and generate the pseudo-ground truth interest points (see Figure 1.6). The resulting detections are more repeatable and fire on a larger set of stimuli.

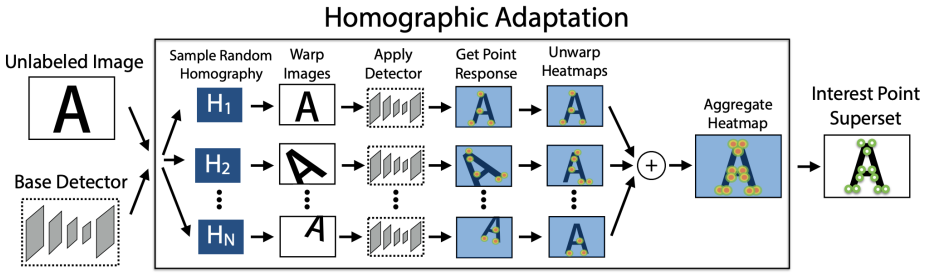

Figure 1.6: **Homographic Adaptation.** Homographic Adaptation is a form of self-supervision for boosting the geometric consistency of an interest point detector trained with convolutional neural networks. (taken from [1]).

Homographic Adaptation technique is applied at training time to improve the generalization ability of the base MagicPoint architecture on real images. The process can be repeated iteratively to continually self-supervise and improve the interest point detector.

1.2.3 SuperPoint Loss Function

The interest point detector loss function L_p is a fully convolutional cross-entropy loss over the cells $x_{hw} \in X$. It is calculated using the following formula:

$$L_p(X, Y) = \frac{1}{H_c W_c} \sum_{h=1,w=1}^{H_c, W_c} l_p(x_{hw}; y_{hw}), \tag{1.1}$$

where $H_c = \frac{H}{8}$ and $W_c = \frac{W}{8}$ for an image sized $H \times W$, Y is the set of corresponding ground-truth interest point labels, individual entries are shown by y_{hw}, and

$$l_p(x_{hw}; y) = -\log\left(\frac{\exp\left(x_{hyw}\right)}{\sum_{k=1}^{65} \exp\left(x_{hwk}\right)}\right). \tag{1.2}$$

The descriptor loss is applied to all pairs of descriptor cells, $d_{hw} \in D$ from the first image and $d'_{h'w'} \in D'$ from the second image. The homography-induced correspondence between the (h, w) cell and the (h', w') cell can be written as follows:

$$s_{hwh'w'} = \begin{cases} 1, & \text{if } ||\widehat{H_{p_{hw}}} - p_{h'w'}|| \leq 8 \\ 0, & \text{otherwise} \end{cases} \tag{1.3}$$

where p_{hw} denotes the location of the center pixel in the (h, w) cell, and $\widehat{H_{p_{hw}}}$ denotes multiplying the cell location p_{hw} by the homography H and dividing by the last coordinate, as is usually done when transforming between Euclidean and homogeneous coordinates. We denote the entire set of correspondences for a pair of images with S.

Finally, the descriptor loss is defined as:

$$L_d(D, D', S) = \frac{1}{(H_c W_c)^2} \sum_{h=1,w=1}^{H_c W_c} \sum_{h'=1,w'=1}^{H_c W_c} l_d(d_{hw} d'_{h'w'}; s_{hwh'w'}),$$

(1.4)

where

$$l_d(d, d'; s) = \lambda_d * s * \max(0, m_p - d^T d') + (1 - s) * \max(0, d^T d' - m_n). \quad (1.5)$$

λ_d is a weighting term used to help balancing the fact that there are more negative correspondences than positive ones, and m_p and m_n are respectively the positive and negative margins of the hinge loss.
The final loss is calculated as the sum the detector loss, and the descriptor loss using the following formula:

$$L(X, X', D, D'; Y, Y', S) = L_p(X, Y) + L_p(X', Y') + \lambda L_d(D, D', S), \quad (1.6)$$

where λ is a factor used to balance the two losses.

1.3 SuperPoint's Performance

When trained on the MS-COCO generic image dataset using Homographic Adaptation, SuperPoint is able to repeatedly detect a much richer set of interest points than the initial pre-adapted deep model and any other traditional corner detector. However, like all other state-of-the-art feature detectors and descriptors, when working with radially distorted images, SuperPoint's performance is not desirable.

Chapter 2

RD-SuperPoint

2.1 Radial Distortion

In geometric optics, distortion is a deviation from rectilinear projection; a projection in which straight lines in a scene remain straight in an image. It is a form of optical aberration. Although distortion can be irregular or follow many patterns, the most commonly encountered distortions are radially symmetric, or approximately so, arising from the symmetry of a photographic lens. These radial distortions can usually be classified as either barrel distortions or pincushion distortions (see Figure 2.1).

Figure 2.1: **Radial Distortion.**

There are different formulas that model radial distortion including the division model [17] and the Brown-Conrady's even-order polynomial model [18]. In this work we have used the division model to distort images with the following formula:

$$x_u = \frac{x_d - x_c}{1 + \lambda r^2} + x_c, \tag{2.1}$$

$$y_u = \frac{y_d - y_c}{1 + \lambda r^2} + y_c. \tag{2.2}$$

where (x_u, y_u) is the pixel coordinate in the original (undistorted) image, (x_d, y_d) is the pixel coordinate in the distorted image, (x_c, y_c) is the distortion center, λ is the distortion coefficient and $r = \sqrt{(x_d - x_c)^2 + (y_d - y_c)^2}$.
We also calculated the inverse of this formula which, given the pixel coordinates of the undistorted image, finds its corresponding coordinate in the distorted image according to the distortion center and coefficient used in the distortion process:

$$x_d = \frac{1 - \sqrt{1 - 4\lambda||x_u - x_c||^2}}{2\lambda||x_u - x_c||^2}(x_u - x_c) + x_c, \tag{2.3}$$

$$y_d = \frac{1 - \sqrt{1 - 4\lambda||y_u - y_c||^2}}{2\lambda||y_u - y_c||^2}(y_u - y_c) + y_c. \tag{2.4}$$

We use this formulas 2.1 and 2.2 to distort images and 2.3 and 2.4 to undistort them. Figure 2.2 shows an example of applying our distortion and undistortion functions over an image from MS-COCO dataset.

Figure 2.2: **Distortion and Undistortion**. The middle image is the result of applying the distortion function over the image on the left hand side. On the right the same image is shown after applying our undistortion function over the distorted image. Note than after distortion some of the image data is lost.

2.2 RD-Invariant SuperPoint

To make SuperPoint invariant to radial distortion we need to distort images before training and combine the homographic adaptation with RD-adaptation when exporting the detections. These modifications add more robustness to radial distortion while keeping SuperPoint's invariance to other image transformations. Figure 2.3 visualizes the comparison between RD-SuperPoint and SuperPoint interest point detector for a radially distorted image from MS-COCO dataset.

2.2.1 Training with Presence of Radial Distortion

In the pre-training phase, SuperPoint applies random homographies on the sysnthetic shapes and their keypoints. In our modification, we distort these warped images and their corresponding keypoints and pre-train the model over distorted images. Figure 2.4 shows synthetic shapes with their keypoints after applying random homographies and distortion.
In addition to the pre-training phase, we also distort the images before the self-labeling phase and generate the pseudo-ground truth for distorted images. Figure 2.5 shows an image from MS-COCO dataset before and after applying random homography and distortion.

2.2.2 Homographic-RD Adaptation

To combine homographic adaptation with RD adaptation we need to add two other layers to the pipeline shown in Figure 1.6. One for distorting the unlabelled images before applying the detector and one for undistorting labeled images after the detector is applied (see Figure 2.6).
After applying the homographies to the original image, we distort them with random distortion center and factor. The distortion center should have its coordinates

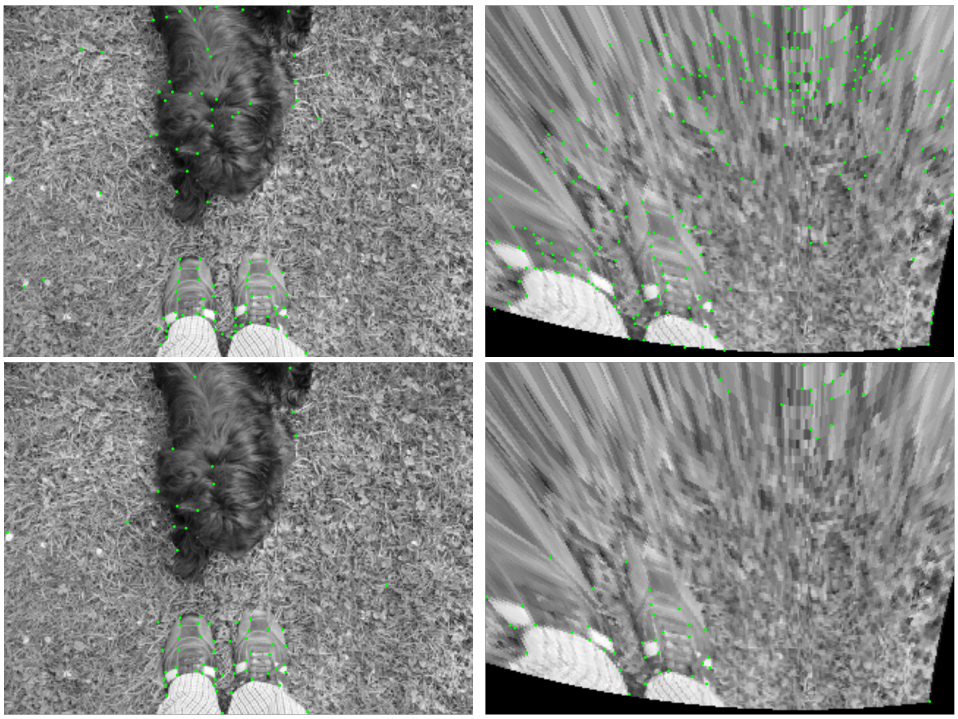

Figure 2.3: **SuperPoint vs RD-SuperPoint Detector**. The upper image shows SuperPoint's detections and the lower image shows RD-SuperPoint's detections. In a radially distorted image, SuperPoint detects many false interest points whereas RD-SuperPoint results is a much lower number of false detections.

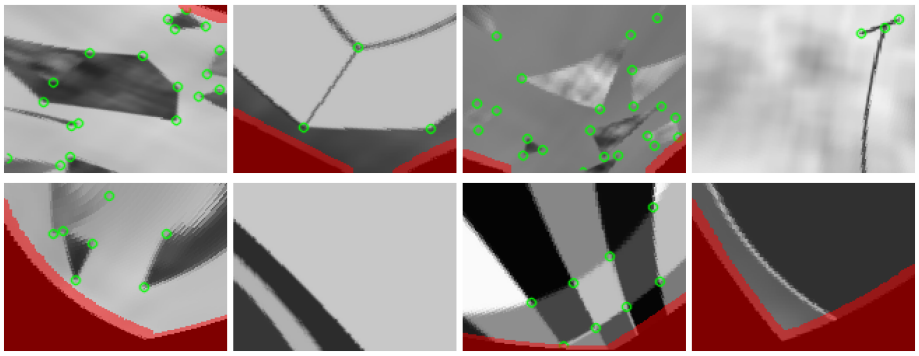

Figure 2.4: **Distorted synthetic shapes.**

within the image and the distortion factor should be small enough to get a reasonable image and large enough to visibly modify the image. For a detailed explanation of proper range for the distortion factors please refer to the "Implementation Details" section.

Once the detector is applied over the distorted images we undistort the labeled images. To do so, we need to undistort the image and the detected keypoints before applying the inverse of sample homographies.

As in SuperPoint, the homographic-RD adaptation can be iteratively repeated to improve the interest point detector. Figure 2.7 visualizes the improvement of our detector after the second iteration.

Figure 2.5: **Combination of Homographic and Distortion Warping**. Image before and after applying homography and distortion.

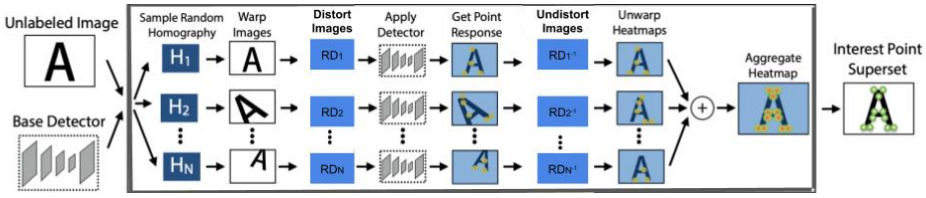

Figure 2.6: **Radial Distortion Adaptation**. By combining homographic adaptation with radial distortion adaptation, we can improve SuperPoint's performance in the presence of radial distortion.

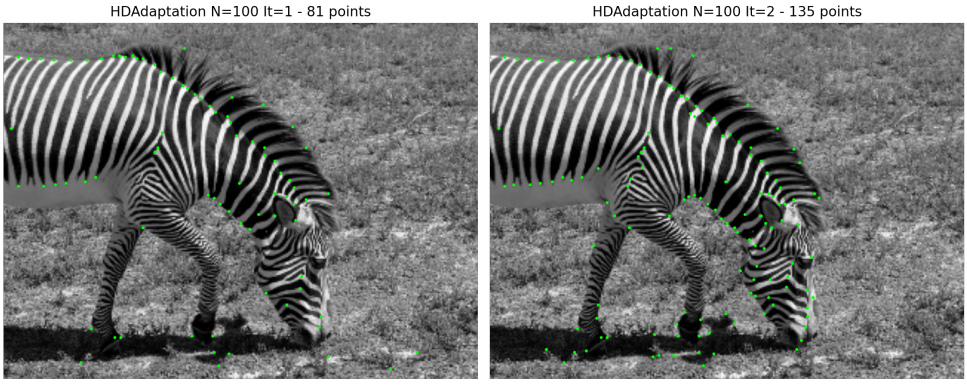

Figure 2.7: **Iterative Homographic-RD Adaptation**. The interest point detector can be improved by repeating the adaptation iteratively.

2.3 Implementation Details

In our implementation, all input images are resize to have width of 320 pixels and height of 240 pixels. As a result all random distortion centers are (x, y) pairs where x is random numbers in $[0, 320]$ and y is a random number in $[0, 240]$.
The distortion factor needs to be large enough to result in visible modification of the original image, and small enough to give reasonable distorted images. After trying several distortion factors, we came to the conclusion that $[0.000002, 0.000009]$ is a proper range for this purpose.
In all the experiments the model is trained on MS-COCO training dataset.

Chapter 3

Evaluation

We evaluate our interest point detector and descriptor both in the presence of radial distortion and its absence. Detailed explanation of each evaluation metric can be found in Appendix A. All evaluations are done using MS-COCO dataset.

3.1 Detector Evaluation

To evaluate our interest point detector, we use repeatability as the evaluation metric. In our evaluations, the repeatability is computed between pairs of images with a maximum of 300 shared points per image.
Table 3.1 shows the repeatability rate for RD-SuperPoint, SuperPoint, and the classical detectors including Harris, FAST, and Shi on images with and without radial distortion.

	With RD	Without RD
RD-SuperPoint	**0.461**	0.748
SuperPoint	0.437	0.680
Harris	0.457	**0.819**
Fast	0.414	0.692
Shi	0.376	0.684

Table 3.1: **Repeatability Rate**. The best rates are highlighted in bold.

As shown in table 3.1, Harris has the highest repeatability rate in the absence of radial distortion, whereas RD-SuperPoint is the most repeatable method in the presence of radial distortion.

3.2 Descriptor Evaluation

We use two metrics to evaluate our interest point descriptor when there is no radial distortion in the images: **Homography estimation**, and **Mean Average Precision (mAP)**. We compute the homography estimation with a correct distance equal 3, and the mean average precision with correctness thresholds from 1 to 30. The matching score is computed with a correctness threshold of 3. Descriptor evaluations are done using a small subset of the MS-COCO dataset containing only 50 images, and the low values of evaluation metrices can be explained by the small size of the evaluation dataset. What is mainly intended to be shown here is that in the same evaluation condition, RD-SuperPoint has a more discriminating descriptor than SuperPoint both in the presence and absence of radial distortion.

Table 3.2 summarizes the performance of the SuperPoint and RD-SuperPoint descriptor when there is no distortion in the images.

	Homography Estimation	mAP	Matching Score
RD-SuperPoint	0.39	0.27	0.29
SuperPoint	0.37	0.24	0.25
SIFT	0.91	0.32	0.31

Table 3.2: **Descriptor Evaluation**. Evaluation in absence of radial distortion.

It can be observed that RD-SuperPoint descriptor has better performance than SuperPoint even in the absence of radial distortion.
When images are distorted, homography estimation is no longer a proper evaluation metric. Thus, in the presence of radial distortion, the descriptor is evaluated based on its mean average precision and matching score. Table 3.3 shows how our descriptor is compared to SuperPoint's descriptor in the presence of radial distortion.

	mAP	Matching Score
RD-SuperPoint	0.26	0.26
SuperPoint	0.24	0.22
SIFT		

Table 3.3: **Descriptor Evaluation**. Evaluation in presence of radial distortion.

Table 3.3 shows that our modifications have enhanced the descriptor's performance. This improvement is expected to be higher in the presence of radial distortion. However, we believe that the dataset used in our evaluations is not large enough to show this improvement.

Conclusion

Radial distortion is a non-linear image transformation present in many images which is caused by the We introduced RD-SuperPoint, a radial distortion invariant features detector and descriptor, which is based on SuperPoint, a state-of-the-art interest point detector and descriptor. Our experiments show that by combining the homographic adaptation technique with RD-adaptation, and generating the pseudo-ground truth for radially distorted images, we can improve the detector's repeatability rate as well as the matching score of the descriptor both for distorted images and images without distortion. These modifications provide a real-time interest point detector and descriptor even is the presence of radial distortion in the images without adding considerable complication to the original method and while preserving all the original invariance of SuperPoint.

Acknowledgement: It is noteworthy that throughout this project a great deal of support and assistance was received by the co-supervisors of this thesis, Rémi Pautrat and Dr. Viktor Larsson. Moreover, the supervisor of this project, Prof. Marc Pollefeys, has played an undeniable role in creating the opportunity of conducting this research is the the Computer Vision and Geometry Group.

Bibliography

[1] T. Malisiewicz, D. DeTone, and A. Rabinovich, "SuperPoint: Self-Supervised Interest Point Detection and Description," 2018.

[2] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," 2006.

[3] D. G. Lowe, "Distinctive image features from scale invariant keypoints," 2004.

[4] M. Calonder, V. Lepetit, M. Özuysal, T. Trzcinski, C. Strecha, and al., "BRIEF: Computing a Local Binary Descriptor Very Fast," 2012.

[5] H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: Speeded Up Robust Features," 2006.

[6] K. M. Yi, E. Trulls, V. Lepetit, and P. Fua, "LIFT: Learned Invariant Feature Transform," 2016.

[7] M. Dusmanu, I. Rocco1, T. Pajdla, M. Pollefeys, J. Sivic, A. Torii, and T. Sattler, "A Trainable CNN for Joint Description and Detection of Local Features," 2019.

[8] J. Revaud, P. Weinzaepfel, C. D. Souza, N. Pion, G. Csurka, Y. Cabon, and M. Humenberger, "Repeatable and Reliable Detector and Descriptor," 2019.

[9] Z. Luo, L. Zhou, X. Bai, H. Chen, J. Zhang, Y. Yao, S. Li, T. Fang, and L. Quan, "ASLFeat: Learning Local Features of Accurate Shape and Localization," 2020.

[10] X. Du, H. Li, and Y. Zhu, "Camera lens radial distortion correction using two-view projective invariants," 2011.

[11] X. Liu and S.Fang, "Correcting large lens radial distortion using epipolar constraint," 2014.

[12] Q. Wang, Z. Wang, and T. Smith, "Radial distortion correction in a vision system," 2016.

[13] M. Lourenco, J. P. Barreto, and A. Malti, "Feature Detection and Matching in Images with Radial Distortion," 2010.

[14] S. Urban and S. Hinz, "MDBRIEF - A fast online adaptable, distorted binary descriptor for real-time applications using calibrated wide-angle or fisheye cameras," 2016.

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014.

[16] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," 2015.

[17] A. Fitzgibbon, "Simultaneous linear estimation of multiple view geometry and lens distortion," 2001.

[18] F. Bukhari and M. N. Dailey, "Automatic Radial Distortion Estimation from a Single Image," 2013.

Appendix A

A.1 Repeatability

The repeatability rate is computed on pairs of images using the following formula:

$$Rep = \frac{1}{N_1 + N_2}(\sum_i Corr(x_i) + \sum_j Corr(x_j)), \tag{A.1}$$

where N_1 and N_2 are the number of points in the first and second image respectively and the correctness is defined as

$$Corr(x_i) = (\min_{j \in 1...N_2} ||x_i - \widehat{x_j}||) \leq \epsilon, \tag{A.2}$$

with ϵ being the correct distance threshold between the two points.

A.2 Homography Estimation

Homography estimation is used to measure the ability of the descriptor to estimate the homography relating a pair of images. To measure the homography estimation of a descriptor we compare how the estimated homography warps the four corners of one image onto the other. Assume H and $\widehat{H}$ are respectively the ground truth and the estimated homographies, and c_1, c_2, c_3, c_4 are the corners of the first image. Then the ground truth corners c'_1, c'_2, c'_3, c'_4 can be obtained by applying H over the corners. To get the estimated corners $\widehat{c}'_1$, $\widehat{c}'_2$, $\widehat{c}'_3$, $\widehat{c}'_4$ we apply $\widehat{H}$ over the corners of the first image. Now the homography estimation can be calculated using the following formula:

$$CorrH = \frac{1}{N}\sum_{i=1}^{N}((\frac{1}{4}\sum_{j=1}^{4}||c'_{ij} - \widehat{c}'_{ij}||) \leq \epsilon). \tag{A.3}$$

A.3 Mean Average Precision

Mean average precision is used to evaluate how discriminating the descriptor is. To calculate this metric we need to find the Area Under Curve (AUC) of the Precision-Recall curve for multiple correctness thresholds using the Nearest Neighbor matching strategy. This metric is computed symmetrically across the pair of images and averaged.

A.4 Matching Score

Matching score evaluated the performance of detector and descriptor. It is computed by measuring the ratio of ground truth correspondences that can be recovered by the whole pipeline (true positives) over the number of features proposed by the pipeline in the shared viewpoint region. Like the mean average precision, matching score is also computed symmetrically across the pair of images and averaged.