

Two steep multi-spectral registration via keypoint detector and gradient similarity

Jehan-Antoine VAYSSADE¹

^aAgrosup D2A2E pole GestAd equipe agriculture de precision 21000 Dijon, France

Abstract

In precision agriculture it's common to use multi-spectral camera (RGB+NIR). Today the necessity of having high spatial resolution and numerous spectral band (6,8,+), have bring new type of cameras such as multi-sensor one, in order to compute spectral index and extract informations such as shape, texture, ... for plants and leaves analysis. The problem with these cameras is that all the sensors are not aligned, so we need a solution to align the different images with precision in close view.

In this study we propose a two step method applied to Airphen camera (6 spectral images). (i) Affine correction using pre-calibrated matrix at different height, the closest transformation can be selected via internal GPS. And (ii) Perspective correction to refine the previous one, using keypoints matching between enhanced gradients of each spectral bands. The contributions of this paper are the evaluation of different types of keypoints detection and their benchmark.

TODO : evaluate keypoints extractors with different parameters

Keywords: Registration, Multi-spectral imagery, Precision farming, Feature descriptor

Email address: jehan-antoine.vayssade@inra.fr (Jehan-Antoine VAYSSADE)

1. Introduction

Image registration is the process of transforming different images of one scene into the same coordinate system. The spatial relationships between these images can be rigid (translations and rotations), affine (shears for example), homography, or complex large deformation models (due to the difference of depth between ground and leaves) [3]. The main difficulty is that multi-spectral images have wavelength with high distance between each spectral bands. Which implies (i) leaves have a different aspect depending on the spectral bands (ii) there are highly complex and self-similar structures in our images (iii) the scene are a grassland or agriculture image at different scale, which is a complex spectral scene making a hard fit for such a registration.

There is two types of registration, feature based and intensity based [12]. (i) Feature based methods use feature matching, in most cases a brute-force matching is used, making those techniques slow. Fortunately these features can be filtered to reduce the matching cost depending of the spatial properties we have, and a GPGPU implementation can reduce the comparisons' cost. (ii) Intensity-based automatic image registration is an iterative process, and the metrics used are sensitive to determine the numbers of iteration, making such method even worth in time for precise registration. Furthermore in multi-spectral we need different metrics for each registered bands which is hard to achieve.

Different studies of images alignment using multi-sensors camera exist using UAV. Some show good performances for feature based [1, 11] with strong enhancement of feature descriptor for matching performances. Other don't and prefer to use intensity based [2] with better convergence metrics, which is slower and not necessarily robust against light variabilities.

Unless this type of articles, as we know, no studies have been made under agricultural and external conditions in near field of view (less than 10 meter). Those studies mainly propose features matching without large methods comparison of their performance (time/precision), spectral band reference selection, or pre-affine correction depending on the distance. Thus, this study propose the best combination of feature extractor and spectral reference on normalized gradients transformation, using pre-affine registration and matches filtering, evaluated at different spatial resolution.

1.1. Material

The multi-spectral imagery was provided by the six-band multi-spectral camera Airphen ¹. AIRPHEN is a

scientific multi spectral camera developed by agronomists for agricultural applications. It can be embedded in different types of platforms such as UAV, phenotyping robots, etc. AIRPHEN is highly configurable (bands, fields of view), lightweight and compact. It can be operated wireless and combined with complementary thermal infrared channel and high resolution RGB cameras. The camera was configured using 450/570/675/710/730/850 nm with FWHM of 10nm. The focal lens is 8mm. It's raw resolution for each spectral band is 1280x960 px with 12 bit of precision. Finally the camera also provide an internal GPS antenna, that can be used to get the distance from the ground.

1.2. Data

Two datasets were taken at different heights. We have used a "stairville LB-3s lighting stand set" like for positioning the camera at different heights. Due to the size of the chessboard, the limited focus of the camera and the height of the lighting stand set, we have bounded the acquisition height from 1.6 meter to 5 meter with 20cm steep.

The first dataset is for the calibration. A chessboard is taken at different heights, the corresponding data can be found in data/steep-chess/. And the figure 1 shows the chessboard taken at each distance from the ground.

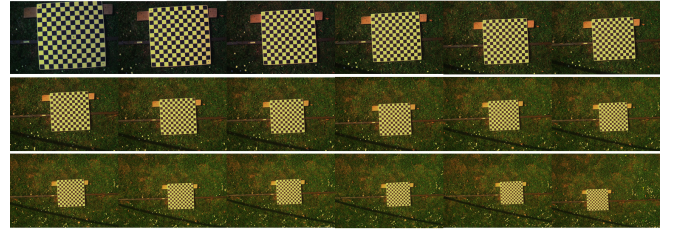


Figure 1: chessboard for calibration at different height

The second dataset is for the alignment verification. One shot of a grassland is taken at different heights, the corresponding data can be found in data/steep/ with a bias of +/- 10cm to be in the worst case (most complex).

¹<https://www.hiphen-plant.com/our-solutions/airphen/>

2. Method

Alignment is refined in two stages, with (i) affine registration roughly estimated and (ii) perspective registration for the refinement and precision. As example the figure 2 show each correction steep at 1.6 meters.

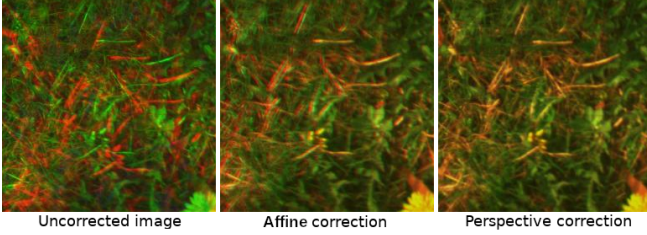


Figure 2: Example of each correction

2.1. Affine Correction

It's important to notice the closer we take the snapshot, the bigger is the distance of the initial Affine Correction. On the other hand at a distance superior or equals to 5 meters, the initial affine correction become stable (figure 3). A calibration is used to build a linear model, which makes the affine correction to work at any height. The main purpose of this step is to reduce the distance of each spectral band, which allow the similarity to be spatially delimited within a few pixels, making feature matching more effective.

Detect chessboard. We use opencv 4 `findChessboardCorners` for each spectral image (normalized) at different heights (from 1.6 to 5 meters). The function attempts to determine whether the input image is a view of the chessboard pattern and locate the internal chessboard corners. The detected coordinates are approximated, and to determine their positions more accurately we use the function `cornerSubPix` as explained in the documentation. The detected points are ordered by x/y (detection can be flipped) and saved on `data/'height'.npy`

Making linear model. Using all the points detected for each spectral band, we calculate the centroid grid (each point average). The affine transform from each spectral band to this centroid grid is estimated. It appear that the rotation and the scale do not depend on the distance to the ground, but only on the translation. This is expected, so that only one calibration can be used for this part of the matrix. This is visible on the figure 3. The factor a, b, c, d is the rotation and scale factor that is quite stable and close to identity (accuracy depends on the spatial resolution of the board). On the other hand, the translation in x, y depend on the height.

Thus a Levenberg-Marquardt curve fitting algorithm with linear least squares regression [5] can be used to fit an equation for each spectral band again x and y independently to the centroid grid. We have chosen to

adjust the following curve $y = ax^3 + bx^2 + cx + d$ where x is the height, y is the resulted translation and factors a, b, c, d are the model parameter.

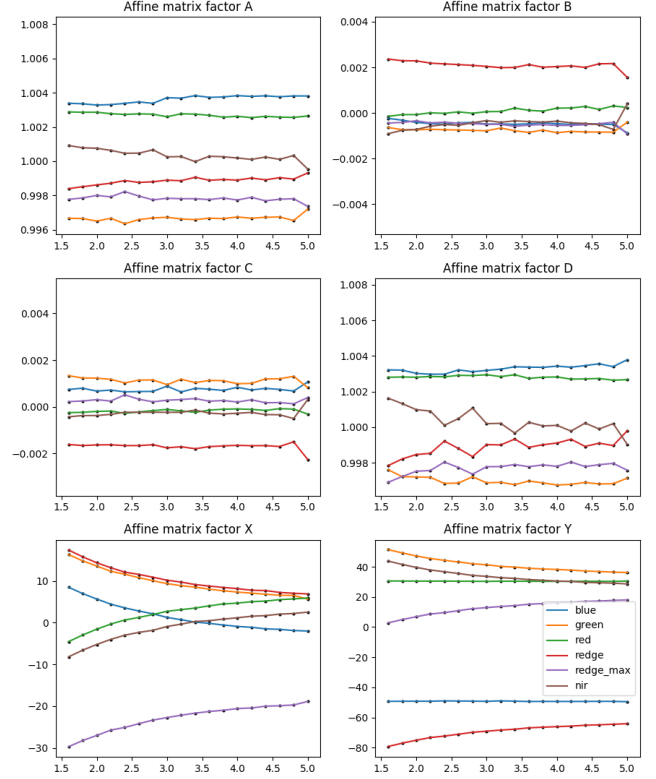


Figure 3: Affine matrix value by height

Correction. To make the affine matrix correction, we used the factors a, b, c, d at the most accurate height (1.6 meter). For the translation part, the curve model is applied for each spectral bands at the given height (roughly known by the user or using the internal GPS sensor). Each spectral band are so warped using the corresponding affine transformation. Finally, all spectral bands are cropped to the minimal bounding box (minimal and maximal translation of each affine matrix).

2.2. Perspective correction

Each spectral band has different properties and values by nature. (figure 4) But we can extract the corresponding similarity by transforming each spectral band into its absolute derivative, to find similarity in gradient break among them.

The previous correction, such as Affine correction attempts to help the feature matching by adding properties of epipolar lines (close). Thus, the correspondence of the extracted features can be spatially bounded, (i) we know that the maximum translation is limited to a distance of a few pixels (less than 10px), and (ii) the angle between the initial element and the matched one is limited to $[-1, 1]$ degree.

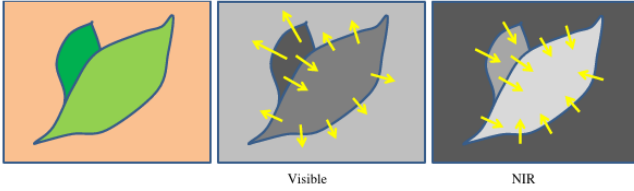


Figure 4: Gradient orientation in spectral band [7]

Computing the gradient. To compute the gradient of the image with a minimal impact of the light distribution (shadow, reflectance, specular, ...) Each spectral band are normalized using Gaussian blur [9], the kernel size is defined by $\text{math.ceil}(\text{image_width}^{0.4}) // 2 * 2 + 1$ (19 in your case) and the final normalized image are defined by $i / (G + 1) * 255$ where i is the spectral band and G are the Gaussian blur of those spectral band. This first step allow to minimize the impact of the noise on the gradient and smoothes the signal in case of high reflectance. Using this normalized image, the gradient are computed with the sum of Sharr filter [10] again $d_x = 1$ and $d_y = 1$.

Different types of edge detection such as Sobel, Laplacian and Canny were tested unsuccessfully (without sufficient number of matches), those not included in this study. Finally, all gradients are normalized using CLAHE [13] to locally improve their intensity and increase the number of key points detected (especially for 850nm).

Keypoints Extractor. A key point is a point of interest. It defines what is important and distinctive in an image. Different types of key point extractors were tested, all the results can be found in "figures/*". These algorithms are all available and easily usable in OpenCV. For all these algorithms, we use the default settings. In some cases, the parameters are set to increase or decrease the number of key points (trying to have a minimum of 20 corresponding key points, and less than 500)..

- ORB : An efficient alternative to SIFT or SURF
 - [1] nfeatures=5000
 - [2] nfeatures=10000
 - [3] nfeatures=15000
- AKAZE : Fast explicit diffusion for accelerated features in nonlinear scale spaces
 - [1] nOctaves=1, nOctaveLayers=1
 - [2] nOctaves=2, nOctaveLayers=1
 - [3] nOctaves=2, nOctaveLayers=2
- KAZE : A novel multi-scale 2D feature detection and description algorithm in nonlinear scale spaces [6]
 - [1] nOctaves=4, nOctaveLayers=2
 - [2] nOctaves=4, nOctaveLayers=4
 - [3] nOctaves=2, nOctaveLayers=4

- BRISK : Binary robust invariant scalable keypoints.
 - [1] nOctaves=0, patternScale=.1
 - [2] nOctaves=1, patternScale=.1
 - [3] nOctaves=2, patternScale=.1
- AGAST : Adaptive and generic corner detection based on the accelerated segment test
 - [1] threshold=71, nonmaxSuppression=True
 - [2] threshold=92, nonmaxSuppression=True
 - [3] threshold=163, nonmaxSuppression=True
- MSER : maximally stable extremal regions
- SURF : Speed-Up Robust Features
 - [1] nOctaves=1, nOctaveLayers=1
 - [2] nOctaves=2, nOctaveLayers=1
 - [3] nOctaves=2, nOctaveLayers=2
- FAST : FAST Algorithm for Corner Detection
 - [1] threshold=71, nonmaxSuppression=True
 - [2] threshold=92, nonmaxSuppression=True
 - [3] threshold=163, nonmaxSuppression=True
- GFTT : Good Features to Track
 - [1] maxCorners=5000,useHarrisDetector=True
 - [2] maxCorners=10000,useHarrisDetector=True
 - [3] maxCorners=15000,useHarrisDetector=True

Keypoint detection. We use one of the keypoint extractors mentioned above between each spectral band gradient (all extractors are evaluated). For each detected keypoint, we extract a descriptor using ORB features. We match all detected keypoints to a reference spectral band (all bands are evaluated). All matches are filtered (distance, position, angle) to eliminate false positives along the epipolar line. Finally we use the function *findHomography* between the key points detected/filtered with RANSAC, to determine the best subset of matches to calculate the perspective correction.

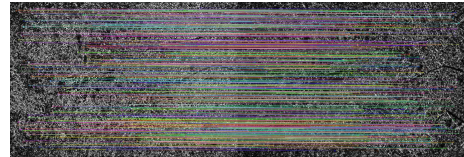


Figure 5: feature matching

Correction. The perspective correction between each spectral band to the reference is estimated and applied. Finally, all spectral bands are cropped to the minimum bbox, the minimum and maximum points are obtained by applying a perspective transformation to each corner of the image.

3. Result and discussion

Affine correction. After the first correction, i.e. the Affine transformation using the linear model from the approximate height, the remaining distance between each spectral band varies according to the distance between the real height and the nearest selected. These residual distances can be seen in the figure 6.

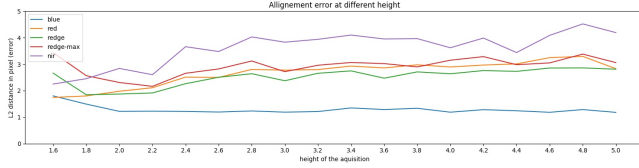


Figure 6: The mean distance of detected keypoint before perspective correction

It appear that the resulted rotation and scale do not depend on the ground distance, but only on the translation. Which expected, so single calibration can be used for this part of the matrix. For the translation part, it depends on the distance to the field, and can be estimated using the fft correlation [8]. Due to the hard correlation between spectral bands these registration, especially between 450nm and 710-850nm (unless using normalized gradient) have not been investigated and suggest that the reader should see the specific article [7].

Keypoint matching. The following figure 7 shows the numbers of keypoints after filtering and homography association (minimum of all matches), the computation time and performance ratio (matches/time) for each method.

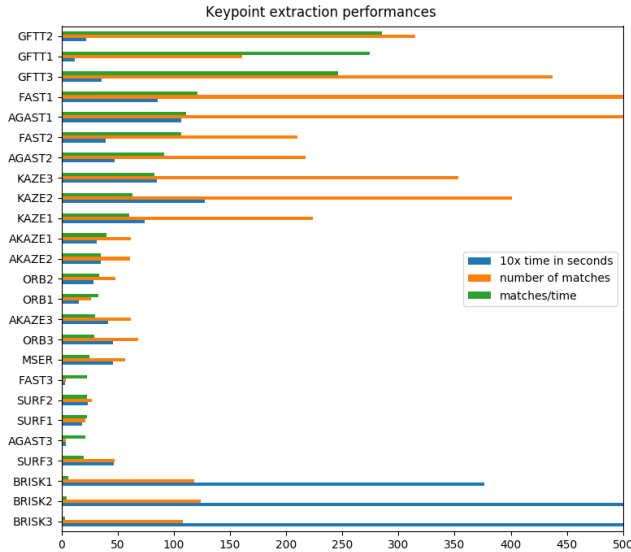


Figure 7: features performances

All these methods work, the choice of method depends on how we want to balance between computation time

and accuracy:

- GFTT show the best performance over all others both in computation time and number of matches
- FAST and AGAST is the most suitable, balanced between time and matches performances.
- KAZE show the best number of matches (>200) but it's also 2.5 times slower than FAST/AGAST.

The other ones did not show improvement in term of performances or matches:

- SURF the number of detected feature may not be enough to fit the perspective correction.
- AKAZE and MSER did not show benefits comparing to FAST.
- ORB could be excluded, the number of matches is near to 20 how is the minimal to ensure that the homography is correct.
- BRISK show good number of matches, but there computation time is too huge (79 sec) comparing to FAST (8 sec).

Increasing the number of key points matched allows a slightly higher accuracy. For example, switch from SURF (30 results) to FAST (130 results). show the final residual distances reduced from 1.2px to 0.9px but the calculation time from 5sec to 8sec.

All methods show that the best reference spectrum is 710nm, with the exception of SURF and GFTT which is 570nm. The following figure 8 shows the minimum number of matches between each reference spectrum and all others using the FAST algorithm. Choosing the right spectral reference is important, as we can see, no correspondence is found in some cases between 650nm-850nm, but correspondences are found between 675nm-710nm and 710nm-850nm, making the 710nm more appropriate, the same behaviour can be observed for the other bands and 570nm. This is visible on the "all_min" line, 570nm and 710nm have the best minimum number of matches. Other best spectral reference are available in supplementary material.

Perspective correction. Residues of the perspective correction show that we have correctly registered each spectral band with a residual error of less than 1 pixel, the figure 9 shows the residual distance at different ground distances.

The following figure 10 showd the difference between detected pointd for two bands (red-green) before (left) and after (right) the perspective correction, and show that the residual errors are spatially uniform.

The decomposition of the residual distance by angle [0 – 45 – 90 – 135 – 135 – 135 – 135 – 180 – 280 – 225 –

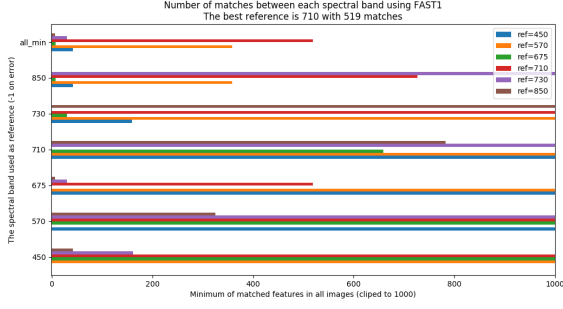


Figure 8: feature FAST performances

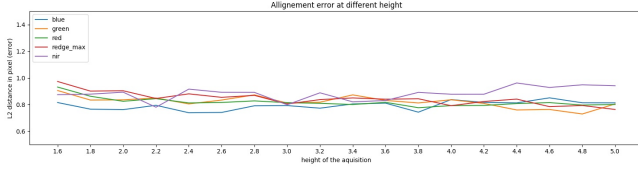


Figure 9: Perspective Re-projection Error

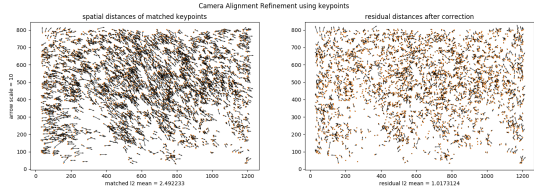


Figure 10: perspective-features-matching-scatter

225 – 225 – 270 – 315 – 0] visible in the figure 11 is interesting. We can notice that the spatial distribution of the residues, for each different angle, is equally distributed. Our hypothesis is that the nature of the base information (spectral band + different lens) makes a small difference to the gradient break, which is detected by the features detector and propagated until the final correction (observed residue). This is interesting because these residues uniformly distributed by angle in space tend to minimize the resulting correction of its center (gradient), thus the detected residual error are overrated and should be less than 0.4 pixel.

However, more complex deformation models [?] can be used to improve the remaining error. This type of complex deformation has not been fully evaluated, but only quickly tested *cv2.ThinPlateSplineSplineShapeTransformer*. There does not seem to be any significant improvement in most cases (with a huge computation time). But can also, in some cases, create large angular deformations caused by the proximity of keypoints, of course, it is possible to filter these keypoints, which also reduces the overall accuracy.

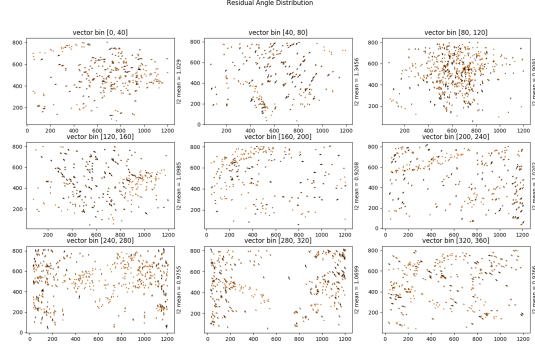


Figure 11: Residual Distribution Again Angle

4. Conclusion

In this work, the application of different techniques for multispectral image registration was explored. We tested different methods of keypoints extractors at different heights and the number of control points obtained. As seen in the method, the most suitable method is the GFTT with a significant number of matches and a reasonable calculation time.

Furthermore, the best spectral reference was defined for each method, for example 570 for GFTT. According to the figure 9 we observe a residual error of less than 1 px, supposedly caused by the difference in input (spectral range, lens). Finally, the method was tested on more than 8000 images in real conditions (not present in the study), randomly taken between 1.6 and 2.2 meters without registration error (always a minimum number of matches, without visible error, less than 0.9px).

Further researches can be performed on each parameter of the feature extractors, for those who need specific performance (time/precision). Otherwise feature matching can be optimized, at this stage, we use brute-force matching with post filtering, but a different implementation that fulfill our epipolar line properties should greatly enhance the number of matches by reducing false positives.

5. Acknowledgment

We would like to thanks Jones Gawain, Combaluzier Quentin, Michin Nicolas and Savi Romain for the realization of the “lighting stand set” that help us for positioning the camera at different heights.

6. Supplementary material

The additional data and source code associated with this article can be found in the online version at the following address <https://gitlab.com/phd-thesis-advent2019/phd-airphen-alignment> thz access is limited, and we invite you to send an email to the author for full access.

7. Reference

- [1] Dantas Dias Junior, J., Backes, A., and Escarpinati, M. (2019). Detection of control points for uav-multispectral sensed data registration through the combining of feature descriptors. pages 444–451.
- [2] Douarre, C., Crispim-Junior, C. F., Gelibert, A., Tougne, L., and Rousseau, D. (2019). A strategy for multimodal canopy images registration. In *7th International Workshop on Image Analysis Methods in the Plant Sciences*, Lyon, France.
- [3] Kamoun, E. (2019). Image registration: From sift to deep learning.
- [4] Lombaert, H., Grady, L., Pennec, X., Ayache, N., and Cheriet, F. (2012). Spectral demons – image registration via global spectral correspondence. In Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., and Schmid, C., editors, *Computer Vision – ECCV 2012*, pages 30–44, Berlin, Heidelberg. Springer Berlin Heidelberg.
- [5] Moré, J. J. (1978). The levenberg-marquardt algorithm: Implementation and theory. In Watson, G., editor, *Numerical Analysis*, volume 630 of *Lecture Notes in Mathematics*, pages 105–116. Springer Berlin Heidelberg.
- [6] Ordonez, A., Arguello, F., and Heras, D. B. (2018). Alignment of hyperspectral images using kaze features. *Remote Sensing*, 10(5).
- [7] Rabatel, G. and Labbe, S. (2016). Registration of visible and near infrared unmanned aerialvehicle images based on Fourier-Mellin transform. *Precision Agriculture*, 17(5):564–587.
- [8] Reddy, B. S. and Chatterji, B. N. (1996). An fft-based technique for translation, rotation, and scale-invariant image registration. *IEEE Transactions on Image Processing*, 5(8):1266–1271.
- [9] Sage, D. and Unser, M. (2003). Teaching image-processing programming in java. *IEEE Signal Processing Magazine*, 20(6):43–52. Using “Student-Friendly” ImageJ as a Pedagogical Tool.
- [10] Seitz, H. (2010). Contributions to the minimum linear arrangement problem.
- [11] Vakalopoulou, M. and Karantzalos, K. (2014). Automatic descriptor-based co-registration of frame hyperspectral data. *Remote Sensing*, 6.
- [12] Zitová, B. and Flusser, J. (2003). Image registration methods: A survey. *Image and Vision Computing*, 21:977–1000.
- [13] Zuiderveld, K. (1994). Contrast limited adaptive histogram equalization. In *Graphics gems IV*, pages 474–485. Academic Press Professional, Inc.