Pixel-wise Hybrid Image Registration on Wood Decors

M. Grunwald, J. Müller, M. Schall, P. Laube, G. Umlauf and M.O. Franz

Institute for Optical Systems, University of Applied Sciences Konstanz, Germany

***Abstract*—The detection of differences between images of a printed reference and a reprinted wood decor often requires an initial image registration step. Depending on the digitalization method, the reprint will be displaced and rotated with respect to the reference. The aim of registration is to match the images as precisely as possible. In our approach, images are first matched globally by extracting feature points from both images and finding corresponding point pairs using the RANSAC algorithm. From these correspondences, we compute a global projective transformation between both images. In order to get a pixel-wise registration, we train a learning machine on the point correspondences found by RANSAC. The learning algorithm (in our case Gaussian process regression) is used to nonlinearly interpolate between the feature points which results in a high precision mage registration method on wood decors.**

I. Introduction

Today a large proportion of all furniture or floors with a wood-like appearance are made of artificially printed wood decors instead of actual wood. Camera-based inspection is used to ensure that the printed decors do not differ from an initial reference. Based on the production environments, line scan cameras are often used for the digitalization of the printed decors. Many defect detection algorithms require an optimal registration of the print and the reference before comparison. This paper describes a registration method for images of wood decors with an accuracy of at least one pixel.

There are several difficulties for this registration process which partly result from the specific image structure of the wood decors and from the setup of the line scan camera. One problem is the repetitive structure of wood. This characteristic leads to a typical correspondence problem for this use case, i.e. image regions at different locations are erroneously matched due to their high visual similarity. Another characteristic of wood is that its visual structure is often oblong and thin—there are many edges but few corners to extract. Edges alone can only be used to locally match image regions in the direction perpendicular to the edge. Along the edge, no unique correspondence between image regions can be established. This is the aperture problem in optic flow processing [1].

Line scan cameras raise more difficulties. Either the line scan camera moves over the image that is digitalized or the printed decor moves on a transport system under the fixed line scan camera. Variations in the speed of movement cause partial stretching or compression of the image in both setups. The difficulty is that this type of image transformation will be different all over the image so that this has to be corrected locally. Similar local distortions arise also because of lens distortions in the line camera.

This paper addresses the problem of pixel-wise image registration on wood decors based on a hybrid registration method. The registration method is built in five steps. The first four steps are a part of the *global image registration* solution which includes the extraction of feature patches, the correlation of these patches, the calculation of model parameters and the validation of the quality. The global registration step can only correct for perspective transformations (including translations and rotations). The fifth step of the registration method can be considered as a *local registration* as local parts of the image are transformed differently to account for movement variations and lens distortion. This step is done with the help of a machine learning method, Gaussian process regression, which leads to a dense, pixel-wise correspondence between both images. This application of machine learning to image registration constitutes the novelty of this paper as this – to our knowledge – has not been done in the literature before.

The paper is organized as follows: we briefly discuss previous work in image registration on wood decors in Section II. In Section III, we present an overview of the global registration and describe the approach on local registration using Gaussian processes. A detailed description and experimental evaluation of each registration step are presented in Section VI. The paper concludes with a discussion in Section V.

II. Previous Work

Image registration strategies are divided into two main classes, both of which deal with different problems. The first class is global registration: an entire image is registered at once by finding its transformation parameters such as translation, rotation, scaling and shearing. In this case, every pixel of an image is transformed in the same way in order to match the other image [2]. The second class of registration problems is the registration of local image regions. In order to register these kinds of images, different transformations for different parts of the images are needed. Within these two classes, there are several subclasses which treat different registration problems [3]. Whereas there is a large literature on image registration in general, the specific problems arising in the registration of wood decors have not been addressed in detail.

In previous work, we addressed the problem of local image registration on wood decors using the classical Lucas-Kanade algorithm for optical flow [4]. This algorithm is widely used to detect movements between pictures or within videos. The application of an optical flow algorithm to the registration process was also proposed by J.-P. Thirion [5]. If an optical flow algorithm is applied to two images which are not registered, it detects the transformations as independent movements of parts of the image. When applied to wood decors, the Lucas-Kanade algorithm ran into problems as the contrast of the wood decor image is not strong enough for the algorithm to detect the optical flow in all parts of the image. In these areas, no transformation could be predicted. Additionally, due to linear structure of the wood grain, the aperture problem affects the vast majority of all image regions so that in these regions only the flow component normal to the edges can be calculated. In order to solve these problems, we applied a variant of the famous *Horn and Schunck algorithm* [6] in which the diffusion of the flow vectors is weighted depending on a contrast-dependent confidence measure. While this method was capable of capturing the local variations caused by movement variations and lens distortion, it performed poorly on estimating the global transformation between the images due to the systematic underestimation of image displacements caused by propagating local displacement vectors affected by the aperture problem.

III. Registration approach

A. Feature selection

The geometric properties of the images are represented by extracted features. Two feature algorithms were examined and compared in this work. Features should have several qualities: (1) they need a strong invariance against small transformations; (2) they have to be localizable which means that the same features are found in both images and that it is possible to match these features; (3) it must be guaranteed that enough features can be found. In an ideal case, these features are evenly spread across the image.

The first feature algorithm examined was the Harris corner detector [7]. The advantage of the Harris corner detector is that it is easy to implement, efficient and that it finds corners independently of their orientation. A recursive Gaussian-like low pass filter was used as pre-smoothing method. For the calculation we used a four point central difference derivative operator. From these filters, the Harris detector computes a local "cornerness" function fo each pixel. A point is considered as a feature point when its "cornerness" exceeds a certain threshold. The threshold has to be adjusted according the image and the number of feature points needed. In order to achieve this, an algorithm was developed which is shown in Fig. 1. To obtain a faster convergence, the desired count of feature points is given in terms of an upper and a lower bound. The step width is only adjusted if the number of features is outside of these bounds.

Harris features have the disadvantage that they are only computed at one image scale and that the corners are found only with a maximum precision of one pixel. They also tend to be very sparsely distributed in the image. We therefore tested a second, scale invariant feature detector: the scale

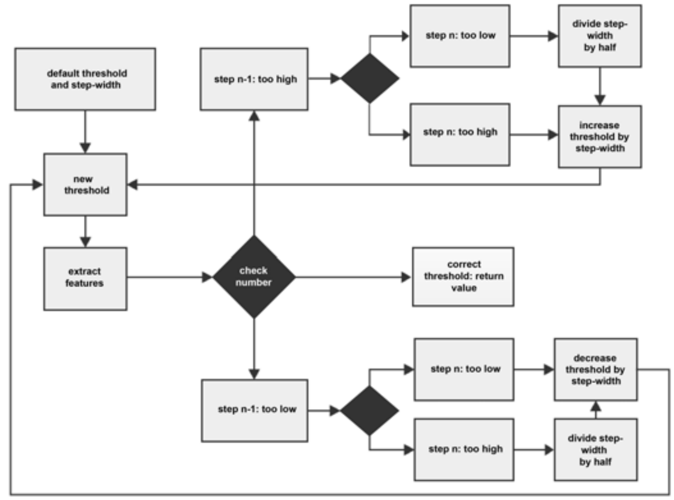


Fig. 1. Algorithm for determining the detection threshold of a feature. n is the number of iterations, the black rhombus symbolizes a decision based on counting the number of the extracted features.

invariant Laplace operator [8], referred to as blob features. The blob operator creates a Gaussian scale space, which is subsampled by a factor of two after every octave. Subsequently a Laplacian scale space is formed by subtracting adjacent layers of the Gaussian scale space. Blob-like image structures are identified as maxima or minima in the Laplacian scale space. The position of blob features can be calculated at sub-pixel accuracy. The underlying theory of the scale space and the blob detector can be found in T. Lindeberg, "Feature detection with automatic scale selection" [8].

B. Global registration

For the global registration we used the RANSAC (*Random Sample Consensus*) algorithm [9] based on the perspective transformation model [10]. The perspective transformation of an image is described by eight parameters a_i, b_i with $i = 1, 2, 3$ and c_j with $j = 1, 2$. That is why at least four corresponding control points in the reference and the transformed image are needed. For each of the point pairs (with index i), the image coordinates (x_i, y_i) in the reference are connected to (x'_i, y'_i) in the print by the equations

$$x'_i a_1 + y'_i a_2 + a_3 - x_i x'_i c_1 - x_i y'_i c_2 = x_i \quad (1)$$

$$x'_i b_1 + y'_i b_2 + b_3 - y_i x'_i c_1 - y_i y'_i c_2 = y_i. \quad (2)$$

These equations are linear in the unknown transformation parameters, so they can be solved by a standard least squares approach. We used the Moore-Penrose pseudoinverse for this purpose. However, the solution requires establishing point correspondences between both images which is the objective of RANSAC. Here, one chooses a large number of randomly chosen subsets of feature points in both images as candidate correspondences and tests the performance of the found transformation on other subsets of feature points. The best performing transformation is chosen to globally register the images. This procedure converges very slow for the large number of features we detect in the images due to its stochastic nature. We therefore replace the random selection of subsets by a more directed form of selection: we extract a local image region around all feature points and find the most similar image region around a feature in the other image by searching for the candidate with the highest Pearson cross-correlation coefficient. All corresponding point pairs found in this way were ranked according to their cross-correlation. We restricted RANSAC to select its subsets only from the group of the highest ranking corresponding point pairs. This crucial step led to a considerable runtime improvement which made the proposed registration procedure feasible at all.

C. Local registration

After applying the global transformation to register both images we applied a second, local registration based on machine learning. The point correspondences between features were used as training data for a nonlinear regression technique. The two-dimensional image positions of the features in the transformed reprint were used as inputs and the x- or y-position of the correspondences in the reference were used as outputs. The result of the training is a dense mapping from 2d positions in the reprint to 2d positions in the reference which can visualized as a vector field (see Fig. 5). In other words, the machine learning interpolates the displacement field between both images at all pixels in the images, not only at the feature points and thus leads to a pixel-wise registration.

We choose Gaussian processes as our regression technique because they adapt well to non-linear functions with added noise, as described by Rasmussen and Williams [11]. This is made possible by the ability to use a covariance function in function-space. In addition, Gaussian processes include only a small number of hyperparameters that can be optimized using gradient descent.

Our implementation of the Gaussian process regression is based on kernel functions [12]. Kernel functions are used to calculate covariance measures in high dimensional spaces without actually transforming the input data. The choice of the kernel function influences the regression function and how well it fits the sample data. In our work, the shape of the mapping was unknown. We tested the Gaussian kernel which can model smooth displacement fields of arbitrary shape and the inhomogeneous polynomial kernel which restricts the shape of the displacement fields to follow two-dimensional polynomial curves. Details on the training of Gaussian processes can be found in the book by Rasmussen and Williams [11]. To find the hyperparameters of the Gaussian process, we used a gradient descent scheme on a smoothed form of leave-one-out error on the training set (Geissers surrogate predictive log probability) [11].

IV. Experimental results

To demonstrate the effectiveness of the proposed pixel-wise hybrid image registration approach we applied the method on different test sets with known and unknown transformations. In order to get representative results the cross validation method was used. For each test set the feature points were divided into equal random subsets and all but one of them were used for the calculation of the solution. This solution was afterwards applied to the unused subset and the error rate of the result was calculated. The error was measured in terms of the average absolute value of pixel deviation between the transformed and true feature point in the test set.

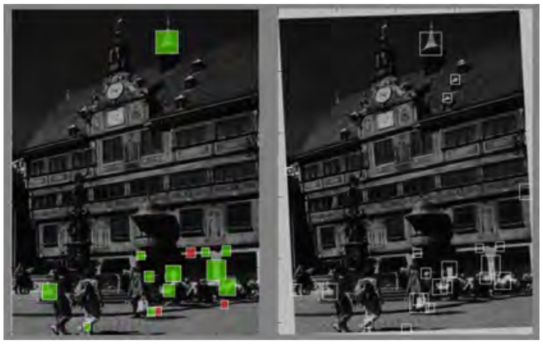
The test sets were divided into different categories. The first test category was artificially generated by applying known transformations. The town hall of the city of Tübingen and an artificially created wood decor served as test images. The contrast to the natural scene is used as an example to demonstrate the specific characteristics of wood decors. The second test category was based on scans of wood decors. Multiple scans of the same decor were made with different scan positions using a line scan camera.

A. Feature extraction

Since feature extraction is important for both the global and the local registration, we first examined the quality of the feature extractors. The term quality is defined as a number of properties of the feature extraction algorithm. The most important property is that the algorithm finds the same features in both images. Thus, it must be resistant to the noise of the camera and the transformations that were applied to the images. The number of features that can be extracted from the image is also of importance. At least four matched features were necessary for the calculation of the perspective transformation. For increasing robustness against imprecise features, ten feature points were used to calculate the perspective transformation by using the pseudo inverse. Additionally, five more features were needed to verify the precision of the calculated parameters. $1/6$ of the features could not be used as they were needed for the cross validation of the whole system. Alltogether, at least about 20 correct matched features were needed for our solution. The precision of the features is a further measure of quality. To achieve the objective of a pixel-wise registration of images, the features need to have a precision of at least one pixel.

The first test was conducted to compare the feature extraction on natural scenes and wood decors. Harris and blob features were extracted from the test images and matched manually to ensure that no mismatch was produced by the correlation algorithm (see Fig. 2). 12 out of 15 features were extracted by the blob algorithm from both images which results in a retrieval ratio of 80%. The Harris algorithm extracted 29 out of 63 from both images, resulting in a retrieval ratio of a about 46%.

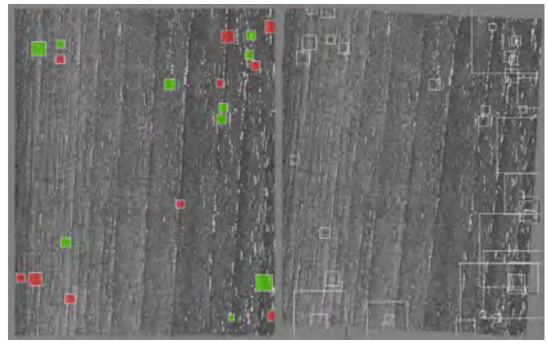
In the second image (artificial wood decor, see Fig. 3), the blob algorithm extracted 10 out of 20 features from both images (50% retrieval rate), the Harris algorithm 12 out of 17 features (70% retrieval rate). Obviously, the quality of the feature detector strongly depends on the texture of the images. Blob features are optimized for images with natural scenes such as landscapes or buildings which is shown by



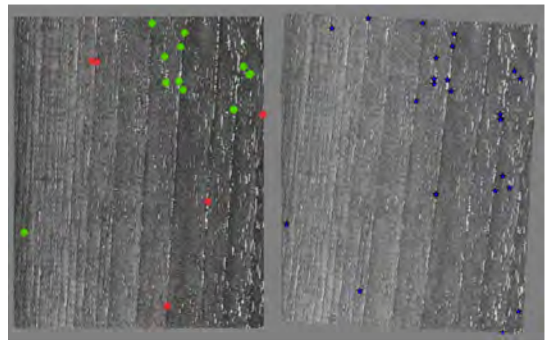
(a)

(b)

Fig. 2. The left images are the originals; in the right images, translation and rotation are applied to the image. The features marked in green can be retrieved in the second image. The red features could only be found in the first image. (a) Correlated blob-features of the town hall of Tübingen. (b) Correlated Harris-features of the town hall of Tübingen.



(a)



(b)

Fig. 3. Same as Fig. 2 for the wood decor.

TABLE I
RESULTS OF THE FEATURE ALGORITHMS ON SCANNED WOOD DECOR.

	Inliers / outliers	**Precision**
Harris, wood1, T1	2.76	0.38
Blob, wood1, T1	1.22	0.18
Harris, wood1, T2	2.15	0.40
Blob, wood1, T2	0.97	0.20
Harris, wood2, T1	1.49	0.28
Blob, wood2, T1	1.06	0.18

their higher retrieval rate for this case, whereas the Harris extractor seems to be more suitable for wood decors. We also measured the average cross-correlation coefficient for both feature algorithms which was much higher for blob features as compared to the Harris features in both images.

In a second experiment, a full registration and a calculation of the ratio of outliers (incorrect correspondences) to inliers was conducted on the scans of the wood decors where the correct transformation is unknown. The results shown in Table I can be seen as indicators of two aspects of quality: the number of common features retrieved from both images and the quality or distinctiveness of the image areas surrounding the feature points.

Table I is structured as follows: the first column describes the test setup: the algorithm, the decor and the transformation used. The number after the transformation in *wood1* indicates two different positions on the same decor during the scanning. The second column contains the quotient of the number of matches (correct matches with a distance smaller than one pixel) divided by the number of mismatched features. As can be seen, the Harris detector always found a larger number of correctly matched features than the blob detector. The third column shows the mean distance of all inliers. This value should be low in order to obtain a high precision. The values of the third column show that the blob features are more precise than the Harris features with a relative improvement of at least 100%. The last aspect of quality is that the algorithm has to be able to extract a sufficient number of features from the image. As we said in the beginning of this section, at least 20 correctly matched features are needed for finding a reliable solution. To guarantee that this can be achieved we found in our experiments that at least 150 feature points have to be extracted from both images. To control the number of features, the threshold for the minimum contrast of a feature was adjusted according to the algorithm described in Fig. 1.

B. Image registration using Gaussian processes

Since this was the first time that Gaussian processes are applied to image registration we first tested whether this method is capable of estimating a known displacement field of a complex shape. Due to its clear horizontal and vertical edges the image of the town hall of Tübingen was chosen for this test. The town hall image was warped two times on a grid, shown in Fig. 4a. No global transformation was applied. The calculation of the Gaussian process response was done by

TABLE II
PRECISION OF THE GAUSSIAN PROCESSES CORRECTION WITH A GAUSSIAN KERNEL.

	[1]without GP	[1]with GP	[2]F-score
Wood1, T1	0.1869	0.1891	3.17
Wood1, T2	0.2189	0.1006	1.17
Wood2, T1	0.1949	0.1210	3.00

[1] Average precision in 5-fold cross-validation.
[2] Number of improved features divided by degraded features.

TABLE III
PRECISION OF THE GAUSSIAN PROCESS CORRECTION WITH AN INHOMOGENEOUS POLYNOMIAL KERNEL.

	[1]without GP	[1]with GP	[2]F-score
Wood1, T1	0.1755	0.1800	2.60
Wood1, T2	0.2259	0.2044	2.07
Wood2, T1	0.1700	0.1570	5.66

[1] Average precision in 5-fold cross-validation.
[2] Number of improved features divided by degraded features.

a set of correlated blob features as trainings points. For this configuration a Gaussian kernel was used.

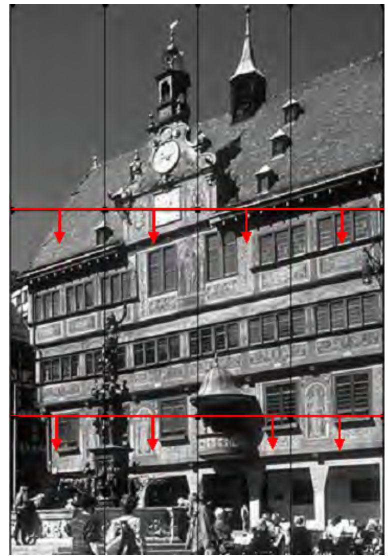
The result of the Gaussian process can be seen in Fig. 4b. The first grid line of the warp in the Gaussian response is where the arrows change the direction. The second warp was done in the same direction as the first line. The resulting arrows, which point in the same direction, follow the correct transformation. The long arrows in the sky above the town hall indicate an incorrectly learned transformation in this region. As the contrast was very low in this area, no features for the training the Gaussian process could be extracted. As a consequence, the predicted displacements vary widly in this area.

C. Hybrid image registration

These tests analyze the performance of the full hybrid image registration method on the scanned wood decor images. First the global registration method were applied to the feature points. In the second step, these corrected features were used to train the Gaussian processes. As a last step, the Gaussian processes were used to predict a correction of unseen feature points in a validation set.

Fig. 5a shows the prediction of the Gaussian processes. The error we observed in the previous experiment in the sky region of Fig. 4b also occurred in the corners in Fig. 5a, where due to the low contrast along the image edges no feature points could be detected which led to a high uncertainty in the prediction of the Gaussian process. The zoomed box shows an enlarged portion of the displacement field. The results for the inhomogeneous polynomial kernel are shown in Fig. 5b. The edge effect observed in the Gaussian kernel is much less pronounced here.

A quantitative evaluation is shown in Table II and III, indicating a significant improvement in accuracy for most cases.



(a)

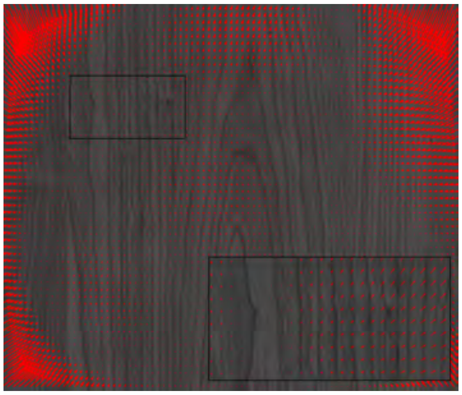
(b)

Fig. 4. Image (a): grid warp of the image. Image (b): response of the Gaussian process.

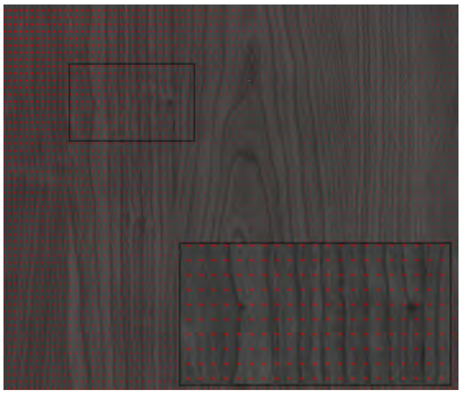
V. DISCUSSION

We presented a method for pixel-wise hybrid image registration on wood decors. Our experiments have shown a considerable improvement in the registration quality using Gaussian processes for local registration.

The results of several experiments show that the blob features achieve a higher accuracy. As a first step, the extracted blob features are correlated by using the Pearson cross-correlation coefficient. To guarantee that the succeeding calculations can be done without outliers, only the feature pairs with the highest correlation are retained. These feature pairs are used as input values for the calculation of the



(a)



(b)

Fig. 5. (a) Prediction of a Gaussian process with a Gaussian kernel on wood decor; (b) prediction of a Gaussian process with an inhomogeneous polynomial kernel on wood decor. Note that the arrows of the displacement field are not to scale, but enlarged for better visibility.

transformation parameters of the projective transformation model using the RANSAC algorithm. To increase the precision of the registration, the local registration is used to interpolate between the already globally corrected feature pairs. Thus, the features are used as training data for the Gaussian processes.

The results of the global and the local registration process depend both on the precision and reliability of the extracted features. The Harris algorithm has better values with regard to the ratio of inliers to outliers, but worse values in terms of the position precision of the features. Since the blob algorithm still has enough inlier features, its higher precision turns out to be more important since the registration should be as precise as possible. For this reason, blob features appear more suitable for registering wood decors according to our experiments. The global registration appears to be very robust against imprecise features or outliers since only the subset with the highest confidence is used to calculate the transformation.

We found a significant improvement in precision due to local registration with Gaussian processes for most cases. In the first test case, however, it is possible that the global registration found a very good solution for the transformation, so that the correction by the Gaussian processes did not improve the precision of the features. The second and third tests results in table II show that the precision is improved – more than 100% in the second test. It comes as a bit of a surprise that the quotient of the second test is the lowest. This might result from the fact that the features which are corrected are rather imprecise after global registration and the correction of those seems to have a strong impact on precision.

The results of the Gaussian kernel and the inhomogeneous kernel are comparable. The predicted correction in both cases is almost completely smooth. It seems that the predictions of the Gaussian kernel vary to higher degree across local image regions and that it is more susceptible to an inhomogeneous feature distribution, whereas the inhomogeneous polynomial kernel generally gives a very smooth prediction. The training data tended to be similar but not equal. The high value of the F-score in the last test in Table III shows that there was still a systematic error in the global registration which the inhomogeneous polynomial kernel corrected best.

As a consequence, in order to use a Gaussian process with a Gaussian kernel, the features should be evenly distributed across the image. There are several options to deal with the areas with a low number of features. One option is to use the predicted variance of the prediction of the Gaussian process which can be calculated together with the predictions [11]. In a low contrast area, the predicted variance would be quite high. In these areas, a correction with a Gaussian process can be disabled or the correction can be done by using the predictions from neighbouring features.

The overall algorithm has not been optimized for real-time application yet. Especially the training process is computationally expensive and needs to be improved to be useful in production systems, e.g. in an inspection system for printed wood decors.

References

[1] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques," *International journal of computer vision*, vol. 12, no. 1, pp. 43–77, 1994.

[2] M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, "Efficient subpixel image registration algorithms," *Optics letters*, vol. 33, no. 2, pp. 156–158, 2008.

[3] B. Zitova and J. Flusser, "Image registration methods: a survey," *Image and vision computing*, vol. 21, no. 11, pp. 977–1000, 2003.

[4] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision." *IJCAI*, vol. 81, pp. 674–679, 1981.

[5] J.-P. Thirion, "Image matching as a diffusion process: an analogy with maxwell's demons," *Medical Image Analysis*, 1998.

[6] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, pp. 185–203, 1981.

[7] C. Harris and M. Stephens, "A combined corner and edge detector." in *Alvey vision conference*, vol. 15. Citeseer, 1988, p. 50.

[8] T. Lindeberg, "Feature detection with automatic scale selection," *Int. J. Comput. Vision*, vol. 30, no. 2, pp. 79–116, 1998.

[9] M. Brown and D. G. Lowe, "Recognising panoramas," *Proceedings Ninth IEEE International Conference on Computer Vision*, pp. 1218–1225 vol.2, 2003.

[10] R. Szeliski, *Image formation*, ser. Texts in Computer Science. Springer London, 2011.

[11] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2006.

[12] B. Scholkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. The MIT Press, 2001.