A proof of concept for aligning sketches to their corresponding painting

Marit Radder
Supervisor(s): Ruben Wiersma, Ricardo Marroquim
EEMCS, Delft University of Technology, The Netherlands

June 19, 2022

A Dissertation Submitted to EEMCS faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering

Abstract

Aligning sketches to their corresponding painting could give more insight into the creative process of an artist. This is a difficult task that cannot be solved directly with classical image registration techniques. Typically, features such as cracks and brushstrokes are used to match different modalities. A sketch does not contain such artifacts and is often roughly similar to the final painting. Therefore the following question rises what suitable feature detection, feature extraction/description, and transform model estimation methods can be used to align sketches to their corresponding painting. This paper provides a proof of concept by taking a manual feature detection, and a histogram of orientations as a feature description. We demonstrate that our algorithm can automatically align sketches with up to 1.4 percent of the accuracy of manual alignment.

1 Introduction

Painters use sketches to prepare for their paintings. For example, Rembrandt is known for using many sketches in exploratory work [1]. Researching those sketches and comparing them to the finished painting can give insight into the creative process of the artist. Moreover, aligning sketches to paintings/drawings could serve more purposes, for example, using it to facilitate a considerable level of ease for 3D modeling [2]. However, manually aligning sketches to painting is time-consuming.

Automatic image alignment has successfully been applied to the painting domain. Conover et al. proposed a method for registering different modalities to the corresponding painting [3]. In this method, craquelure is used to extract the key points. However, craquelure cannot be used for a sketch to painting registration since there is no craquelure in the sketch. Another issue that is present when comparing sketches to paintings is the fact that a painting is not an exact copy of the sketch. For example, the shapes and colors of paintings and sketches differ. Given these limitations, other registration methods need to be developed to successfully register sketches to paintings.

This research paper will answer the questions of whether sketch to painting registration is possible and which feature detection, feature matching, transform model estimation, and which image re-sampling and transformation can be used to register sketches to paintings.

The proposed method described first extracts the edges using the Traditional Inspired Network (TIN) [4]. Then, the user guides the feature extraction step by manually selecting regions of interest. Those features are described using the Gradient Field HoG (GF-HOG) [5]. Next, the features are matched using a brute force matching method. This is followed by calculating the average transform between those matches. This transform is applied to the sketch.

In the following section, the related work is discussed. Then we provide some background on the techniques used in the proposed method. This is followed by an overview of the proposed method. Next, the experiments are explained and the results are discussed. Lastly, the conclusion and some future work are given.

Figure 1: The extracted edges (left) of De Nachtwacht (top right) and its corresponding sketch (bottom right) using TIN

2 Related Work

Image registration Image registration methods can be used for an image to painting registration [6]. Aslan et al. conclude that using feature-based approaches performed poorly due to the significant gap between source and target images and that using features from feature-based methods leads to the detection of too many low-quality interest points. Therefore, a multi-modal intensity-based registration method is used. Other multi-modal registration methods use craquelure to select keypoints [3]. Craquelure is present in the painting, however, it is not present in the sketch. Looking into intensity-based registration methods it should be noted that the intensity of a sketch versus a painting could differ severely. One example is that the contrast in sketches, and therefore the difference in intensity values, are generally higher than in paintings. Considering these differences between sketches and images/paintings a new image registration method needs to be developed.

Feature description Angular partitioning is used to describe features in sketches and images [7]. Angular partitioning the descriptor describes spatial information of the edges [7]. The local spatial structure could also be described, this is done using a Gradient Field HoG (GF-HOG) descriptor. It specifically encodes the spatial orientation of the edge [5]. The contour segments descriptor combines both GF-HoG and angular partitioning into one descriptor [8]. Contrary to angular partitioning and GF-HOG this method uses contours as features instead of edges. Another way to model the features extracted from the edge is by implicit polynomials to provide a similarity computation method that is robust towards user query sketch distortions [9]. Shape similarity is also used to describe features in sketches. Here the edge of the sketch will be warped to fit the edge of the image [10]. Since the local

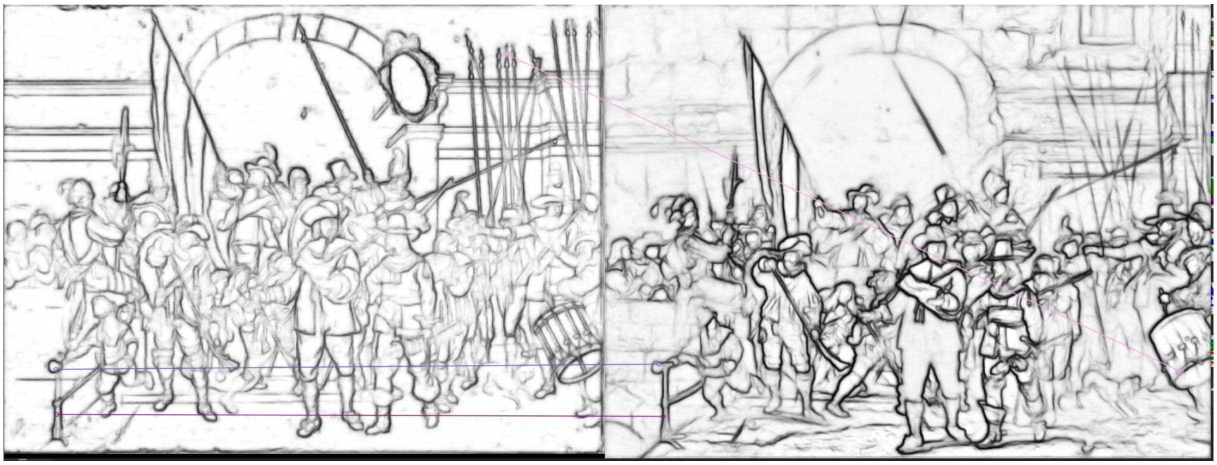

Figure 2: The result of the matching algorithm with 4 regions of interest

spatial information of edges in both the sketch and the painting would roughly be the same, GF-HOG is used.

Feature matching The Nearest Neighbour method is a feature matching method that looks if the neighbouring features of a feature in a sketch are also close to the neighbouring features of the feature in the reference painting [11]. Random sample consensus (RANSAC) is also used for the alignment of the sketch with the reference picture [5]. RANSAC estimates a global relation that will fit the data. It classifies the data into inliers and outliers [12]. In the proposed method the number of features is restricted, therefore a brute force matching method is used.

Image alignment Cross-correlation is used to align different modalities to their corresponding painting [3]. The phase images of both the modality and painting will move over each other and will get convoluted. Then the coordinates of the highest peak will be the translation. Iterative closest point(ICP) could also be used to register free-forms [13]. In ICP the difference between two cloud points is minimized. It will converge to the closest local minima. However, that does not necessarily mean that is the optimal solution. Therefore, the method uses cross-correlation for its transformation.

3 Method

The proposed method consists of the following steps:

- Edge detection
- Feature extraction
- Feature matching for rough alignment
- Cross correlation for detailed alignment

The proposed method will take as input two scans or images, one from a painting and one from the corresponding sketch. The proposed method is not scale-invariant, therefore the content of the images needs to be of the same size.

3.1 Edge detection

The Traditional method Inspired Deep Neural Network (TIN) [4] is used for the edge detection. TIN is a Deep Neural Network that includes a Feature Extractor, Enrichment and a Summarizer which is designed to roughly correspond to the gradient calculation, low pass filter and pixel connection in traditional edge detection. It is used because it showed good results when used in image to painting registration [6]. The result of using TIN on De Nachtwacht can be found in Figure 1

3.2 Feature extraction

The proposed method uses annotations to extract features. Since the edges extracted from the sketches and paintings would still differ, more traditional feature extractors like the Harris Corner method would not extract the same features. By annotating the features there is more control over selecting the same features. The user is asked to select 8 areas of 64 by 64 pixels in both the sketch and the painting. A minimum of one matching region of interest (ROI) is necessary for finding a transformation. Selecting more ROIs would make the algorithm less prone to outliers. However, they will give more overhead and thus make the algorithm more computationally expensive.

Sketches and paintings differ a lot. Paintings are most of the time in colour where a sketch is not. Another example is that a hand in a painting could move a bit compared to the painting. But even the lines themselves differ, the line work of a sketch is less accurate than a painting. If a bigger region of interest would be selected, those little differences would be accounted for. However, selecting a bigger area would also give more overhead and therefore be more computationally expensive.

3.3 Feature description

The proposed method uses Histograms of Gradient of Orientations as the feature descriptor. Because it can handle small positional changes in the region of interest. And it encodes the local spatial information instead of the global spatial information as it would in using angular partitioning. It will take the orientation of the edge in a region of interest and form a histogram. In the proposed method the histogram is divided up into 8 bins. So they all have a range of 45 degrees.

Making a histogram instead of comparing the exact position makes the proposed method less prone to differences in the position of the edges in the sketch and the painting.

3.4 Feature matching and Transformation

The algorithm uses a brute force matching algorithm instead of RANSAC or a nearest neighbour matching algorithm. This is due to the fact that the selected amount of regions of interest are small. The brute force matching algorithm will compare the descriptor to all other descriptor's features and return the feature with the smallest distance. The result of applying it on the edges of De Nachtwacht can be found in Figure 2. Once the regions of interest are matched the transformation matrix will be found by taking the average translation of the center of the matched regions of interest in the X and Y direction.

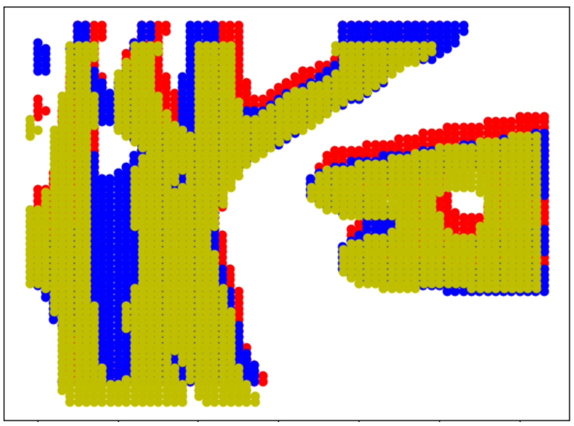

Figure 3: A closer look at the exact translation using cross correlation. In red the sketch, in blue the painting and in yellow the exact translation

3.5 More exact alignment

The alignment using manually selected regions of interest is not as exact as automatically extracting the features. This due to the fact that humans are not as pixel-perfect as an algorithm. Now that the sketch and painting are already roughly aligned, a more exact alignment can take place without making it too computationally expensive. The more exact alignment is done using a phase cross-correlation. The cross-correlation returns the x and y coordinate of the position of the peak, which corresponds to the x and y translation. The result of the more exact translation can be found in Figure 3.

4 Experimental Setup and Results

Several experiments are conducted on the algorithm, to find the most optimal variables and to validate the proof of concept. The algorithm has three variables that need to be optimized. Namely, the number of regions of interest (ROI), the size of the ROI and the number of orientation bins. Lastly, some minor tests will be done to test the whole proof of concept of the algorithm. The experiments are performed on a small test set consisting of 7 sketches and paintings. Two of those will include sketches and paintings made by Rembrandt the rest is a more simplistic version of a sketch and a painting. The test set could be requested.

One variable is changed at a time. The other variables are fixed to the following values: the amount of ROIs is 4. The size of the ROIs is 64 x 64 and the amount of orientation bins is 8.

The results are evaluated by comparing the percentage of overlap between the aligned sketch and painting. This is done by taking the xor between the sketch and the painting in every position divided by the amount of pixels times 100 minus 100. This is compared against a base measurement using a manual alignment. A higher percentage of overlap than a manual alignment does not necessarily mean the algorithm performs better. Since sketches and paintings differ, it could be that the algorithm finds a bigger overlap between two different objects. This would return a higher percentage of overlap, although the alignment is worse. Therefore a good result for this proof of concept would be a percentage of overlap that is close to the percentage overlap of the manual alignment. That would show that the proof of concept performs as good as a manual alignment.

The accuracy of the matching is also evaluated by dividing the sum of True Positives (TP) and True Negatives (TN) by the number of selected regions of interest. Lastly, the precision is also evaluated by dividing the amount of TP by the sum of TP and TN.

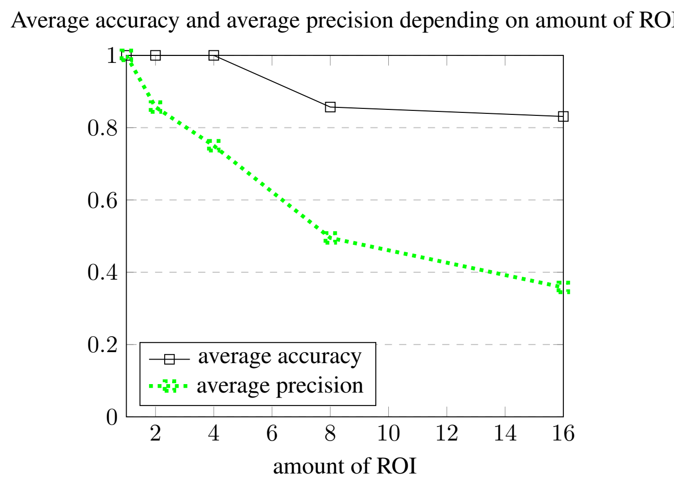

Figure 4: The average accuracy and average precision plotted against the amount of ROI

4.1 Amount of regions of interest

Five different ROI amounts are evaluated: 1, 2, 4, 8 and 16. Since the position of a ROI is a factor in the success of the algorithm, the position of features is the same across the different amounts. Meaning that the first 4 ROIs of the 8 ROIs are the same as the 4 ROIs selected when the amount of ROIs is 4. The ROIs is get manually selected.

By taking the average accuracy and average precision of the test set the plot in Figure 4 is found. While the accuracy is above 80 percent, the precision falls off. Although

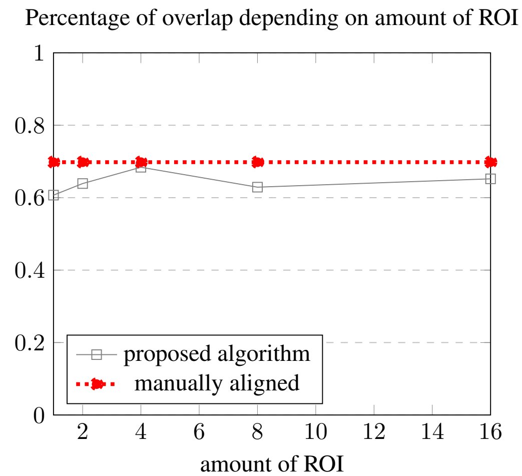

Figure 5: The amount of overlap plotted against the amount of ROI

the biggest problem would be false matches, a lower precision would mean a lower amount of matches. If looking at the number of good matches, it can be seen that that number stays around four. Even though the number of ROI increases.

In Figure 5 the percentage of overlap depending on the amount of ROI is plotted. The figure shows that the average amount of overlap is always lower than when the images are manually aligned. It also shows that the result also not influenced much by the different amounts of ROI.

Based on Figure 4 and 5 four is the amount of ROI that will be used in the algorithm.

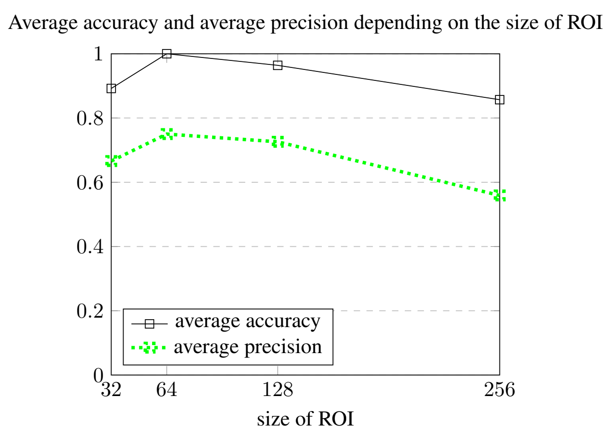

Figure 6: The average accuracy and average precision plotted against the size ROI

4.2 The size of the regions of interest

Four different ROI sizes will be evaluated. These will be the sizes 32 x 32, 64 x 64, 128 x 128 and 256 x 256. Since the position of the ROI is a factor in the matching algorithm. The positions of the ROIs will also be constant in this case. These will have the same position as in the experiment with the amount of ROIs.

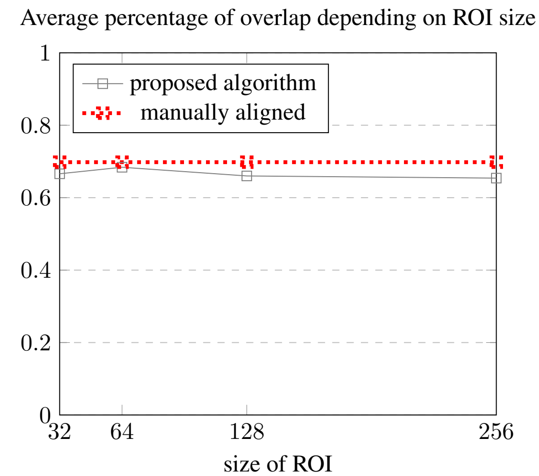

Figure 7: The average overlap plotted against the size ROI

In Figure 6 the average accuracy and average precision depending on the ROI size are plotted. It can be seen that both the accuracy and precision fall off slightly after a ROI size of 64. The size of the ROI does not influence the overlap of the alignment. Based on these results the size of ROI will be 64 pixels by 64 pixels.

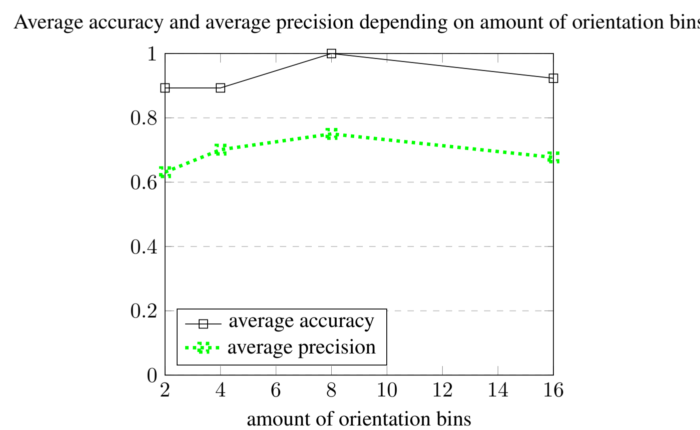

Figure 8: The average accuracy and average precision plotted against the amount of orientation bins

4.3 The number of orientation bins

Four different experiments will be done to evaluate the optimum amount of orientation bins. It will evaluate the following amounts: 2, 4, 8, and 16. Also in these experiments the position of the ROI will be constant.

The average accuracy and average precision against the number of orientation bins can be seen in 8. There is a small optimum around the amount of 8 orientation bins regarding accuracy and precision. The number of orientation bins does not influence the amount of overlap that much, however, it does become close to the manual alignment when more orientation bins are used. However, since the accuracy and pre-

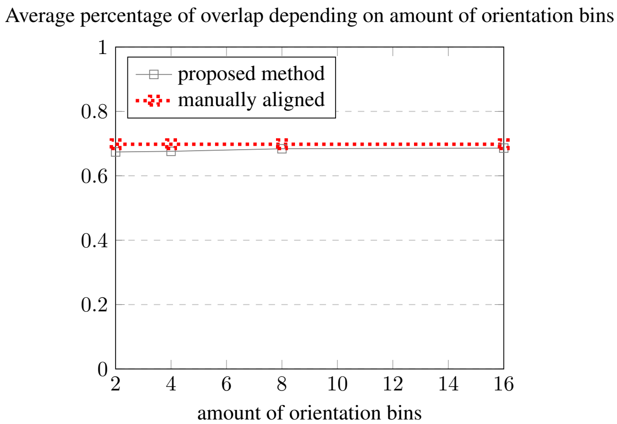

Figure 9: The average overlap plotted against the amount of orientation bins

cision are worse there, it is decided that the number of orientation bins will be 8.

4.4 Cross-correlation

The use of cross-correlation is also evaluated. This is done by comparing the average percentage of overlap before and after the more precise alignment computed by the cross-correlation. The average percentage of overlap with only the rough transformation is 68,4 percent, whereas the amount of overlap with cross-correlation is 67,9 percent.

These differences are small, however, cross-correlation is more computationally expensive. The same or slightly better result can be achieved without the use of cross-correlation, therefore it is decided to leave out the precise transform in the algorithm.

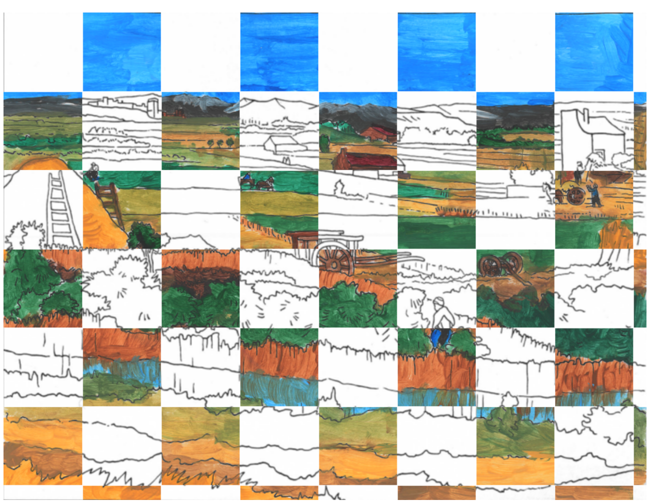

Figure 10: The result of stitching the sketch with the painting

4.5 The pipeline as a whole

Although the percentage of overlap, average accuracy and average precision are metrics that can give insight into how the algorithm performs. It does not necessarily mean that the result is good. A higher overlap does not necessarily mean that the alignment is correct. Therefore some extra tests are executed using the fixed values of a ROI amount of 4, ROI size of 64 and orientation bins amount of 8 and no cross-correlation. In these experiments it is judged if the alignment visually looks good.

In Figure 10 the result of a stitch between a painting and sketch is displayed. Although the stitch seems good at first sight, a further inspection shows a minor misalignment in some parts. This is true in all cases, the alignments are not perfect.

5 Discussion

Sketches, if used in the painting registration field, are now manually aligned to paintings. This is a tedious and time-consuming task. And since there exists no specific algorithm to register sketches to painting yet, it can only be compared to a manual alignment. Looking at Figure 5, 7, 9 it can be seen that the overlap is close to that of a manual alignment. Figure 10 also show that the alignment visually looks good, however not perfect.

Although the algorithm is specifically made for a sketch to painting registration, multiple fields use the same techniques. The sketch-based image retrieval field and the 3D modelling field both use techniques on which the method is based.

It cannot be concluded that the algorithm performs better than a manual alignment only based on the percentage of overlap. A higher overlap does not necessarily mean a better alignment. Even a percentage of overlap that is the same as a manual alignment would not mean it has the same alignment as the manual alignment. Additionally, the definition of a good alignment also depends on the person. The researcher could align the main object to see the difference in the background, or it could align the background to see the difference in the main object. However, it is a sufficient metric to conclude if it would be feasible to get a good result out of the proof of concept.

It could be argued that the percentage of overlap is quite low. However, this is because sketches and paintings differ quite a lot. This could be due to perspective changes, or composition changes. Another factor is that the average translation of the regions of interest is taken. If looked at the percentage of overlap in one region of interest, this is higher. The difference of overlap between the whole image and one region of interest was up to 9,8 percent in some cases.

It is concluded that the use of cross-correlation does not leave a positive significant mark on the result. Even though the difference in the percentage of overlap in the regions of interest can be up to 15,5 percent. However, since an average of all the translations is taken for the final translation to the sketch, the result could be a translation that is not optimal.

6 Responsible Research

The dataset used in the experiments is quite small, seven different painting and sketches pair were used. This is due to the fact that there are not a lot of sketches and painting pairs out there. This in combination with the fact that we had some constraints on the input resulted in a small data set. Efforts were made to create extra sketches and paintings to increase the size of the dataset. However, it should be noted that these sketches and paintings are more simplistic than the paintings and sketches from professional painters such as Rembrandt. These simplistic versions are sufficient for a proof of concept but are not representative for professional artwork.

The result of the algorithm is compared to a manual alignment. However, aligning two things that are not exact copies is a subjective task. One could decide to align the background to see the subject moving, or one could align the subject to see the background moving. The manual alignment was also tested visually instead of using a metric because that is the current way the art conservators do it. Therefore it could be that it would not be the most optimal alignment with the respect to the amount of overlap. However, the transformations of these alignments are stored and can be requested for the reproduction of these experiments.

Annotating the features is also a task which highly influences the experiments. Annotating the features one pixel to the left could give different results. These annotations were not optimized but done in a way to select features that would visibly be distinctive. The coordinates that were used for this data set could also be requested for the reproduction of these experiments. These annotations are done by the creator of the algorithm. To properly test if algorithms like these could replace the manual alignment, a user study should be done to test if the algorithm is user-friendly, but also if an art conservator could get the same results out of the algorithm.

7 Conclusions and Future Work

This paper tries to answer the question of whether sketch to painting registration is possible and which feature detection, feature matching, transform model estimation and image resampling and transformation can be used to register sketches to paintings. The proposed method shows that sketch to painting registration would be possible. This can be done using the deep learned edge detector TIN [4], a manual selection of 4 regions of interest of a size of 64 pixels by 64 pixels, a feature description using GF-HOG using 8 orientation bins, a brute force matching method. The use of cross-correlation for a more precise transform was concluded to be ineffective. Although this is a proof of concept the average percentage of overlap only differs by 1,4 percent.

7.1 Future Work

Although the proposed method performs close to a manual alignment, it does not outperform a manual alignment and it is not fully automatic yet. Certain improvements can be made to improve the proposed method to make this proof of concept into an algorithm that can be used.

Firstly to remove the need for a human interaction a keypoint selection need to be implemented. It would need to select on keypoints both present in the sketch and painting. This could mean that the sketch and painting need to be preprocessed better to highlight those features better. The proposed method now uses a traditional inspired deep learned edge detection. However, the deep learned edge detection Sketch Tokens [14] can also be used. The detector is learned on sketches which could be useful for this use case. If both the processed sketch and painting have the same type of features, a feature detector like Harris Corner could be used. If the keypoint selection would return more key points than what currently is the case, a different matching method is also suggested.

The transformation step of the algorithm also could be greatly improved. Currently, only a translation will happen between the sketch and painting. This could be extended to include a full transformation. In the proposed method it is chosen to apply the transformation to the sketch as a whole, the decision can also be made to apply different transformations to different regions of the sketch. This would improve the amount of overlap between the painting and sketch, however, information about the composition could be lost. Cross-correlation could be used to find these local transformations, as could improve the amount of overlap in local areas.

References

[1] M. Royalton-Kisch, "Rembrandt's sketches for his paintings," *Master Drawings*, vol. 27, no. 2, pp. 128–145, 1989. [Online]. Available: http://www.jstor.org/stable/1553928

[2] C. Li, H. Lee, D. Zhang, and H. Jiang, "Sketch-based 3d modeling by aligning outlines of an image," *Journal of Computational Design and Engineering*, vol. 3, pp. 286–294, 7 2016.

[3] D. M. Conover, J. K. Delaney, and M. H. Loew, "Automatic registration and mosaicking of technical images of old master paintings," *Applied Physics A*, vol. 119, pp. 1567–1575, 6 2015.

[4] J. K. Wibisono and H.-M. Hang, "Traditional method inspired deep neural network for edge detection," in *2020 IEEE International Conference on Image Processing (ICIP)*, 2020, pp. 678–682.

[5] R. Hu, M. Barnard, and J. Collomosse, "Gradient field descriptor for sketch based retrieval and localization." IEEE, 9 2010, pp. 1025–1028. [Online]. Available: https://ieeexplore.ieee.org/document/5649331

[6] S. Aslan and L. Steels, "Identifying centres of interest in paintings using alignment and edge detection," 2021, pp. 589–603. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-68796-0_42

[7] A. Chalechale, G. Naghdy, and A. Mertins, "Sketch-based image matching using angular partitioning," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 35, pp. 28–41, 1 2005. [Online]. Available: https://ieeexplore.ieee.org/document/1369343

[8] Y. Zhang, X. Qian, and X. Tan, "Sketch-based image retrieval using contour segments." IEEE, 10 2015, pp. 1–6. [Online]. Available: http://www.ece.ualberta.ca/~vzhao/temp/Papers/MMSP15_119.pdf

[9] Y. Chans, Z. Lei, D. P. Lopresti, and S.-Y. Kung, "Feature-based approach for image retrieval by sketch,"

C.-C. J. Kuo, S.-F. Chang, and V. N. Gudivada, Eds., 10 1997, pp. 220–231.

[10] R. Mehrotra and J. Gary, "Similar-shape retrieval in shape data management," *Computer*, vol. 28, pp. 57–62, 1995.

[11] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–110, 11 2004.

[12] R. Raguram, J.-M. Frahm, and M. Pollefeys, "A comparative analysis of ransac techniques leading to adaptive real-time random sample consensus," pp. 500–513, 2008.

[13] Z. Zhang, "Iterative point matching for registration of free-form curves and surfaces," *International Journal of Computer Vision*, vol. 13, pp. 119–152, 10 1994.

[14] J. J. Lim, C. L. Zitnick, and P. Dollar, "Sketch tokens: A learned mid-level representation for contour and object detection." IEEE, 6 2013, pp. 3158–3165.