MUSIC SCORE IMAGE REGISTRATION

FOR ANNOTATION EXTRACTION

By Joseph Grealy
joseph.p.grealy@durham.ac.uk

WHAT IS IMAGE REGISTRATION?

Image registration is the process of aligning two or more images of the same scene. These could have been captured at different times, by different devices/sensors, or from different viewpoints.

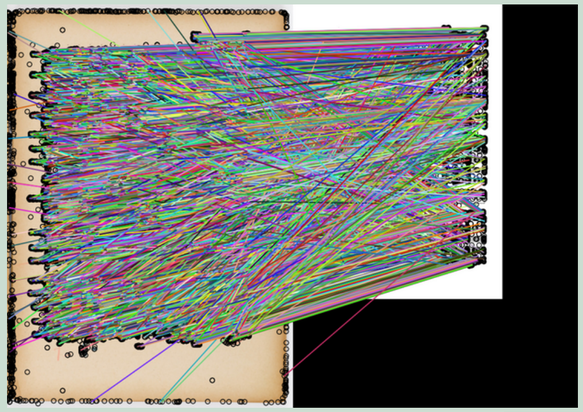
Some examples of image registration are:

- in medical imaging to align different types of scans (like MRI and CT scans).
- in satellite imaging to align and compare changes over time (such as of glaciers).
- in document scanning to correct for skew and other deformations.

There are two main types of image registration:

1. Intensity-based which compares the pixel values of two images to find common patterns.
2. Point-based which finds corresponding points in both the images and uses these to register the image.

RESULTS



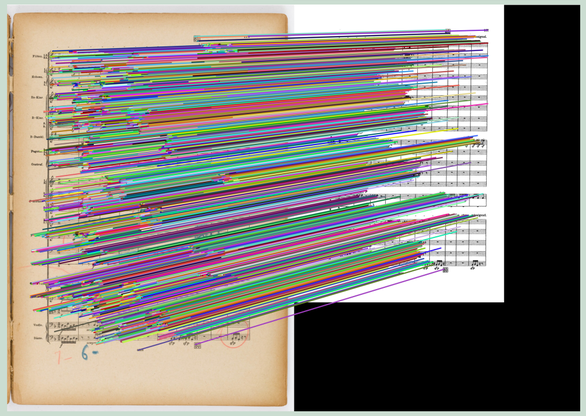


Fig 1: To the upper left shows all of the SIFT keypoints and the initial 5822 matches. It is hard to make out but there are many false matches indicated by the diagonal criss-crossing lines. Correct matches should be almost horizontal between corresponding features in the scores.

Fig 2: To the lower left shows all 2208 matches remaining after outliers were removed by the RANSAC algorithm. These appear to be correct, between corresponding points on the printed score. These matches can then be used to calculate the transformation to register the clean score.

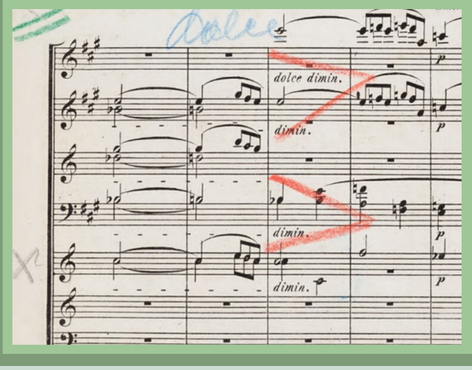
Fig 3. To the right are the images obtained from subtracting the registered clean score from the annotated score, and the same with the binarised (black and white) images. The registration appears largely successful, but binarising reveals lots of noise left over from the printed score, this is due to pixel-differences in the quality of the scans. Light coloured annotations are also not visible due to the thresholding process when binarising.

WHAT ARE MUSIC SCORE ANNOTATIONS?

Musicians enrich musical scores by adding handwritten annotations reflecting their interpretation of the music, and providing additional playing instructions.

The digital extraction of these annotations is of interest to musicians and historians alike.

Being able to digitally access the annotations made by renowned musicians would be use to musicians for playing, and of cultural value.

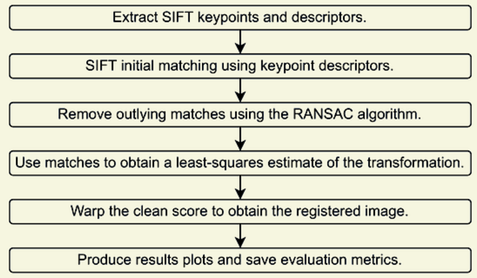


EXTRACTING ANNOTATIONS

If it is possible to obtain a clean, unannotated copy of an annotated score of interest, we could align images of the two scores and "subtract" them leaving behind just the annotations.

Luckily there are digital music score libraries such as IMSLP which contain clean scans of hundreds of thousands of scores.

The annotated scores I used for this project were obtained from the New York Philharmonic Orchestra's digital archives I had 17 image pairs with the clean copies obtained from IMSLP.



METHOD

Since the two images we are registering are not identical (one has annotations), point-based image registration techniques are more applicable since it only relies on parts of the image that are the same!

Most point or feature-based methods follow the same steps:

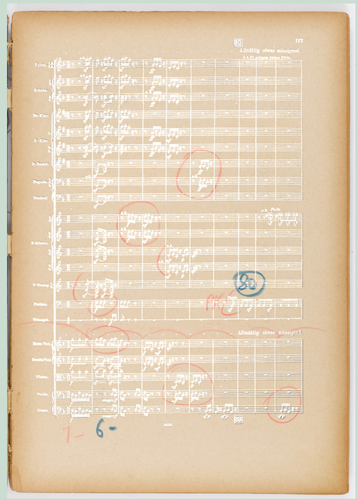
1. Locate feature points in the images.
2. Find matching points between the two images.
3. Use the matches to estimate a transformation that will align the two images.

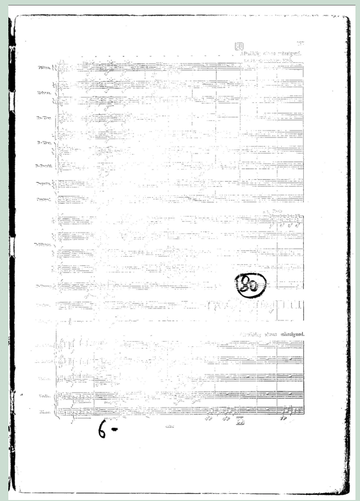
A popular feature extraction method is David Lowe's SIFT (Scale Invariant Feature Transform) algorithm. This locates keypoints and computes highly distinctive keypoint descriptors which are good for matching.

I chose also to use the RANSAC (RANdom SAmple Consensus) algorithm which would remove the majority of incorrect matches returned by the SIFT algorithm.

I also made the assumption that deformations between the two images would be some combination of translation, rotation, global scaling, and shear. This meant I had to find a transformation which could "undo" each of these. Specifically this is a type of linear transform called an affine transform.

The exact steps that were followed are shown in the diagram to the left.





CONCLUSIONS

Upon examining the results, the registration is successful in most cases and handles the spectral differences in the images well. In a number of cases parts of the image are slightly off. This is indicative of perspective warping in the scanning process meaning a different transformation model should be explored.

In addition, the task of actually extracting the annotations from the difference image proved difficult. Advanced denoising techniques could be used to remove the noise left over due to slight differences in the printed scores. Supervised machine learning could also be used "learn" what is annotation and what is noise from ground-truth images where the annotations have been manually labelled.

Research was carried out by Joseph Grealy studying Maths and Statistics under the supervision of Dr Eamonn Bell in the department of Computer Science at Durham University