DOI 10.5162/SMSI2023/P01

Non-Overlap Image Registration

Stefan Siemens[1], Markus Kästner[1], Eduard Reithmeier[1]
[1] Institute of Measurement and Automatic Control, An der Universität 1, 30823 Garbsen, Germany,
stefan.siemens@imr.uni-hannover.de

Summary:

This work aims to predict the relative position of non-overlapping image pairs consisting of a moving and a fixed image. For this purpose, a modified VGG16 convolutional neural network is proposed. The network is trained on a large dataset with microtopographic measurement data of different materials and processing methods. The proposed method shows a high prediction accuracy on the test data and the potential for developing non-overlap registration algorithms.

Keywords: Image Registration, Surface Metrology, Convolutional Neural Networks, Machine Learning, Jigsaw Puzzle Problem

Background

Optical measurement methods are an essential tool in the research and development of technical components. They allow for a fast, accurate, and non-damaging acquisition of the component's micro-topography and its subsequent characterization [1]. Measuring with a high lateral resolution, e. g. being able to resolve small structures, leads to a significant decrease in the field of view and, therefore, the statistical significance of the characterization based on a single measurement. A solution to this problem is given by image registration. It is a widely used tool to find the geometric relation between multiple measurements to unify them into a single coordinate system. Use cases range from consumer tech, like generating panoramic photographs, to metrological use cases. Here they enable capturing surfaces with a high spatial resolution and a large spatial extent simultaneously.

Motivation and Objective

Traditionally, intensity- and feature-based registration algorithms have been used, with the latter still widely used today [2]. However, these traditional approaches are often lacking with today's requirements. In recent years deep learning and convolutional neural networks have achieved impressive results in many disciplines concerning computer vision or audio and text processing. Learning-based approaches have also shown promising results in image registration, where they often outperform traditional methods [3]. However, if the images to be acquired do not have any shared visual content, i.e., if they do not overlap, both classical and learning-based approaches cannot be used [4]. Therefore, this work aims to investigate a method to find the geometric correspondence of an image pair if they do not overlap or even if there is a gap. This problem is simplified by limiting the registration to a classification problem of nine classes, e. g. a center image and its 8-connected neighbors. The problem shows similarities with the jigsaw puzzle problem [5]. However, no inference can be made based on the image's shape. In addition, the images are not directly adjacent to each other but have gaps.

Methods and Data

In order to estimate the relative position of two images, a VGG16 Net with batch normalization is modified and trained in a supervised manner. The VGG Net is a convolutional neural network introduced by Simonyan and Zisserman [6]. The architecture is shown in Fig. 1. The input layer consists of the image pair. It is followed by a feature extractor, e. g. alternating convolutional and pooling layers.

Fig. 1. Architecture of the modified VGG16 Net.

The extractor is followed by a classificator, consisting of three fully connected layers with 25088, 4096, and 9 neurons, respectively. The network outputs the estimated class, e. g. the position of the moving image in relation to the fixed image (the center image). The cross-entropy loss, and the Adam optimizer were used with a batch size of eight. The learning rate was scheduled with an initial value of 0.0001, patience of 5 epochs, and a decay factor of 0.1. Each input pair consist of a moving and a fixed image. The fixed image is always the center image (class

C5), while the moving image is either an image patch from the 8-connected neighbors or the center patch (classes C1-C9). These nine patches are cropped from microtopographic measurements of different surfaces, see Fig. 2.

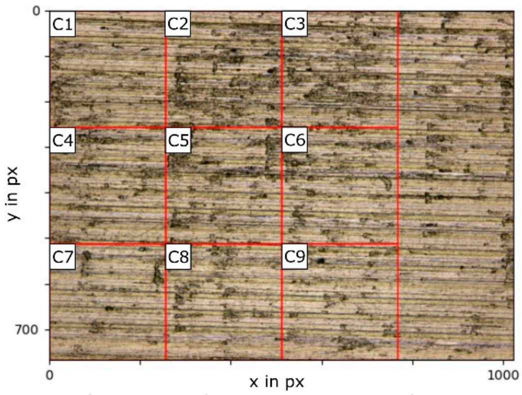

Fig. 2. Cropping of the nine patches from a microscopic image (5x magnification) of a horizontally milled Nickel alloy surface with R_a=6,3 µm and R_z=32 µm.

In total, surfaces with 23 different processing methods were measured. Many processing methods were manufactured with different roughnesses, for example, Ra=0.55 µm, 1.0 µm, 1.6 µm, 3.0 µm, 6.0 µm, and 10 µm for flat ground specimens. Tab. 1 shows five exemplary processing methods from the dataset.

Tab. 1: Examples of materials and processes used in the dataset.

Material	Manufacturing Process
Al_2O_3	Thermal Spraying
Nickel Alloy	Flat lapping
Nickel Alloy	Polishing
Nickel Alloy	Spark erosion
Rubber	Foam

The images were acquired using the two confocal laser scanning microscopes (CLSM) Keyence VK-X210 & Keyence VK-X3000. Objective lenses with 2.5x, 5x, 10x, 20x and 50x magnification were used. The sensors of both microscopes have a resolution of 1024x768 px. Each measurement contains a 2.5D height map, a laser-intensity image, an RGB image, and a combination of the latter two. For this work, combined laser-intensity-RGB images were used. In total, 55.210 measurements were taken. From these measurements, 306.603 patches were cropped and used for training, 101.826 patches for validating, and 102.735 for testing the network. The patches have a size of 256x256 px. From each patch, 240x240 px are randomly cropped, so there is a gap between 0 px and 45 px or 0-18% of the patches' side length.

Results

Fig 3. shows the result after 50 epochs of training. It can be seen that the mean accuracy for the validation set is 97.35% and 96.85 % for the test set, respectively. The highest accuracy was achieved for the center class. This result is in line with the expectation, as there is no movement between the fixed and moving image in this case. The other classes are predicted with an accuracy between 95.48 % and 97.31%.

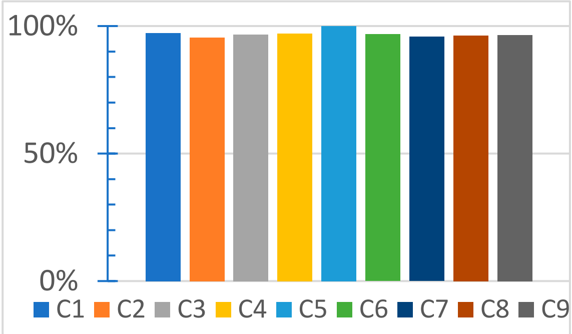

Fig. 3. Percentage of correct predictions on the test dataset for each class.

Conclusion

This work shows that learning-based algorithms can be used to predict the relative position between two images. The predictions have very high accuracy and are robust for many surface types and gaps between the images. The results offer much potential for future work in which, for example, the prediction of homographies for non-overlapping images could be investigated.

References

[1] X. Jiang, et al., Paradigm shifts in surface metrology. Part II, Proceedings of the Royal Society A 463, 2071-2099 (2007); doi: 10.1098/rspa.2007.1873

[2] S. Paul, U.C. Pati, A comprehensive review on remote sensing image registration, International Journal of Remote Sensing 42, 5396-5432 (2021); doi: 10.1080/01431161.2021.1906985

[3] N.J. Tustison et al., Learning image-based spatial transformations via convolutional neural networks: A review, Magnetic Resonance Imaging 64, 142-153 (2019); doi: 10.1016/j.mri.2019.05.037

[4] C. Guérin et al., Detection of comic books twin pages with a non-overlapping stitching method, MANPU 16 (2016); doi: 10.1145/3011549.3011550

[5] S. Markaki, C. Panagiotakis, Jigsaw puzzle solving techniques and applications: a survey, The Visual Computer, (2022); doi: 10.1007/s00371-022-02598-9

[6] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, ICLR 2015 (2015); arXiv:1409.1556v6