2/16/2021

Hyperspectral Image Preprocessing

Image Registration

Student: Rareș-Tudor Buzatu
Coordinator: Ion Necoara
FACULTY OF AUTOMATIC CONTROL AND COMPUTER SCIENCE

Table of Contents

Abstract

This paper has the purpose to present the research done in this semester. In the next pages will be presented the introductory notions regarding both Hyperspectral Images and Image Registration, such as: definitions from both fields, actual problems that need to be solved regarding these subjects, State-of-the-Arts methods that solve these problems and the choice made in selecting one of the methods. All the things achieved by now in the process of solving this topic will be presented, along some following ideas and plans for the following semester. The main problem that is presented here is the problem of aligning two images, in order to get a perfect match between their information and the position of each pixel. The case studies selected represents 2 real world problems: first is image registration on medical images such as CT scans in collaboration with Colatea Hospital , the other is the detection of algae from the North Sea by hyperspectral analysis in colaboration with Norwegian University of Science and Technology (NTNU).

Introduction

For this academic year, I chose to study "Hyperspectral Image Preprocessing: Image Registration". As mentioned before, the coordinator professor of this project is I. Necoara and together we want to improve the current State-of-the-Art methods for image registration in the field of medical imagistic and work on other fields of applications as well.

Context

The reason of choosing this problem is the immense need in the medical field [1] and geographical field of this type of improvement. Having in mind that most medical procedures involve some body scans in order to determine a disease or a condition and most treatments of these conditions are given based on the results of the scans, it is vital to have an accuracy as high as possible in order to treat the patient correctly [1]. At the same time, many geographical scans require seeing through blurred or opaque elements, so another type of vision is required. For example, observing groups of algae in the ocean requires observing things through various environments such as water and air [2].

Problem

One of the many problems of these kind of scans it is represented by the alignment of the images of patients between scans. This problem can lead doctors to misreading the results and to giving patients wrong treatments. The causes of this problem are multiples and they involve different factors such as the position of the patient on the table, the vibrating parts of the machine that executes the scan, the imperfections of the table, the breathing of the patient. Plotted below are two computer tomograph scans that appear to be aligned, but in fact are misaligned. The first image (A) represents a postoperative radiograph of two screws inserted in the cervical region of the spine. The second image (B) represents the first axial CT scan and the third image (C) represents the second axial CT scan of the same patient [1]. In theory, the second and third image should be perfectly aligned, but due to patient movement, the spine seems to be much higher in the first axial scan, than in the second scan [1].

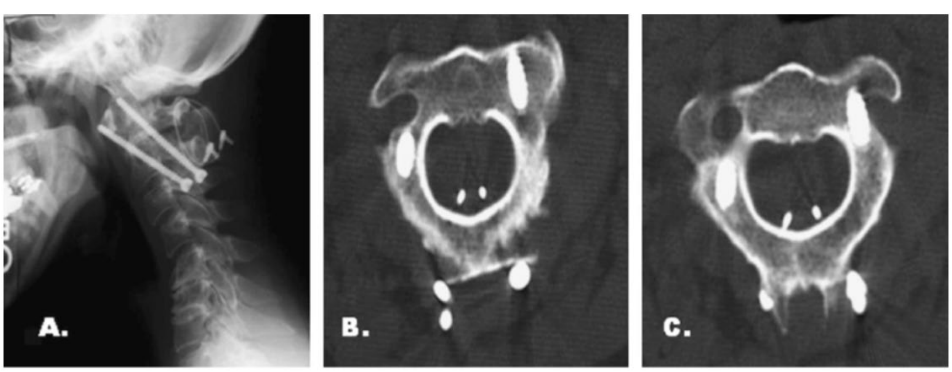


Figure 1: Misaligned CT photos [3]

Regarding the geographical aspects of image registration, images taken from an airplane in order to observe algae are blurred because of the high reflectance of water, thus the way of viewing the algae has to be done differently.

Objectives

The objective of this research is to find a solution that maximizes the similarity between two photos like the ones from above, so the overlap of the objects in them will have a percent closer to 100. Combining information from all the mathematical / computer vision world, medical world and geographical world, the particularization of the State-of-the-Art method of Image Registration should be improved for particularizing the problem to this kind of images. However, the objectives for this semester were to understand these concepts of registration and hyperspectral imaging and to find some solutions that could be improved later on.

Structure of the paper

In the first chapter of this paper some basic image registrations terms and principles will be defined in order to accommodate the reader with the concept, then some techniques for implementing this image pre-process step will be illustrated.

The second chapter will be about hyperspectral images and what they represent. A comparison between hyperspectral images and RGB images will be done and the differences between them will be highlighted. Here, a more detailed overview about the algae problem will be presented.

The third chapter will present the medical need of registration, along with the problems that come along this field.

In the fourth chapter the link between the two fields will be explained and some State-of-the-Arts methods will be described along with the chosen method that was considered to be the best for this task.

In the last chapter, there will be presented the practical experience that I have gained during this semester along some experiments, some results that I came with and the plans for the following semester will be shown.

Image registration

Image registration is a pre-processing algorithm that is applied to images in order to align them, that means that the input of this algorithm has to be at least two images that represent the same object [4]. After the image registration is done, every pixel from the two input images should, in theory, represent the same part of the object. Depending on the input images, the registration can be done taking into account several aspects.

Input Classification

1. Multi-temporal registration [5]

FIGURE 2: MULTI-TEMPORAL REGISTRATION [6]

This is the case where the input images are taken having between them a time interval big enough in order that the object modify its structure / color / shape / texture. For example, from a medical point of view this is the case of MRIs or CTs taken from 6 months apart in order to see the evolution of cancer in a patient, or from a geographical point of view, this is the case of taking 2 images of the same place from the same angle in 2 different seasons of the year.

2. Multi-modal registration [5]

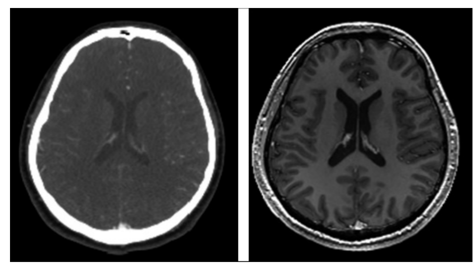


FIGURE 3: MULTI MODAL REGISTRATION*

The multi-modal registration refers to the modality of capturing the images, where different sensors are used in order to capture various aspects of the object that is being analyzed. This is the case of hyperspectral images, where for every wavelength a different sensor is being used and the images are taken one after another, with almost 0-time delay.

3. Multi-view registration [5]

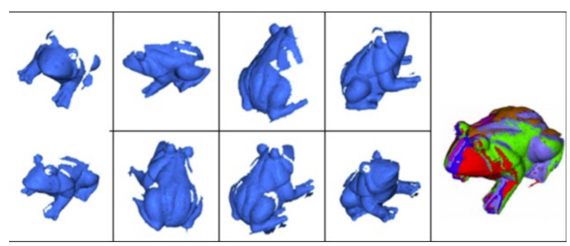


FIGURE 4: MULTI-VIEW REGISTRATION**

The multi-view registration consists of two or multiple images taken from slightly different angles of the same object. This is the case of 3D modelling where 2 or more cameras are used and their results have to be aligned in order to build the 3D model of the object.

4. Combination between the first two

Many applications have more than one entity that is varying, so in the most cases, the registration is a combination of the previous classes.

**D. Jiang, "Fast and robust multimodal image registration using a local derivative pattern", Medical Physics Volume 44 Issue 2, 09 december 2016*

*** Batlle E., Matabosch C., Salvi J. (2007) Summarizing Image/Surface Registration for 6DOF Robot/Camera Pose Estimation. In: Martí J., Benedí J.M., Mendonça A.M., Serrat J. (eds) Pattern Recognition and Image Analysis. IbPRIA 2007. Lecture Notes in Computer Science, vol 4478. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-72849-8_14*

Medical registration is definitely a multi-view process because of the patient constantly moving, a multi modal registration because of the sensors and procedures of most scans and if the registration is between scans that are done several months apart, then is also a multi-temporal registration. On the other side, the algae detection is first a multi-modal registration because of the hyperspectral type of images that are being captured and second, is a multi-view because of the movement of the plane producing changes in the angle of the camera regarding the algae.

Algorithm

The main structure of registration algorithms is the same from their earlier methods to these days. The structure has some typical steps that need to be followed in order to achieve this pre-processing.

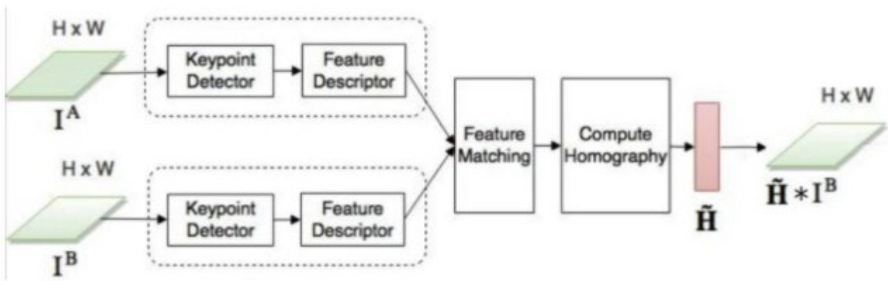


FIGURE 5: IMAGE REGISTRATION PROCESS [7]

Keypoint Detection.

The first and the main important task in the image registration process is to find the correct keypoints. Keypoints are distinctive parts of an image that build its unicity, so in order to detect an object we have to take into account its most distinctive parts. For example, in Fig. 5, a comparison between good keypoints and bad keypoints is made.

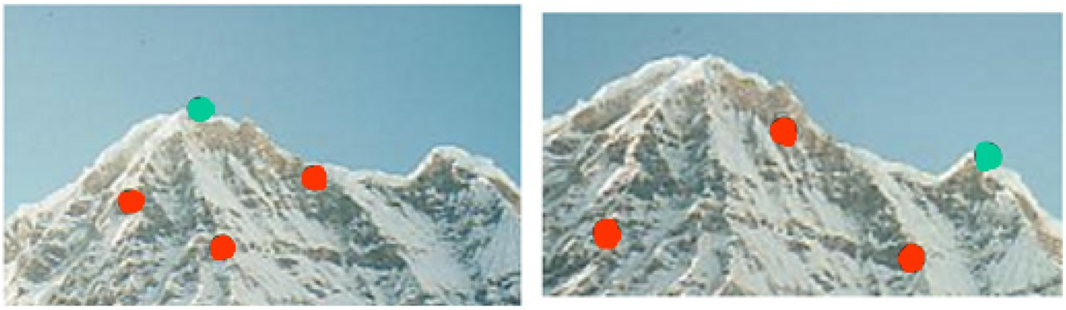


FIGURE 6: KEYPOINTS: GREEN= GOOD ORANGE= BAD [6]

If someone wants to detect not just a mountain, but a particular mountain, that person should chose as keypoints the edges of the mountain because that is what gives that mountain its particularity, not the snow that is on it, or the trails / rocks that it has.

In most cases, the biggest particularity of an object is given by its shape, so it is intuitive to consider the contour of the object as a keypoint area selection. However, the contour is not always the most distinctive part. Other distinctive parts could be the texture or the color. The contour is nothing else but a high derivative of the image, so we can compute it by the formulas [6]:

$$I_x \Leftrightarrow \frac{\partial I}{\partial x} \qquad I_y \Leftrightarrow \frac{\partial I}{\partial y} \qquad I_x I_y \Leftrightarrow \frac{\partial I}{\partial x}\frac{\partial I}{\partial y}$$

FIGURE 7: CONTOUR FORMULAS [6]

The most powerful part of the contour are the corners, or in other words the derivative on both axes. However, in order to identify the objects by this method we have to keep in mind the size of the objects in the image, so multiple scales have to be used when searching for corners. The most used method to do that is to take the Laplacian-of-Gaussian function and to modify its sigma in order to capture different scales. The Laplacian depends on both derivates on X and Y, so the result will be influenced by a round blob of different sizes:

$$\nabla^2 g = \frac{\partial^2 * g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2}$$

Laplacian-of-Gaussian

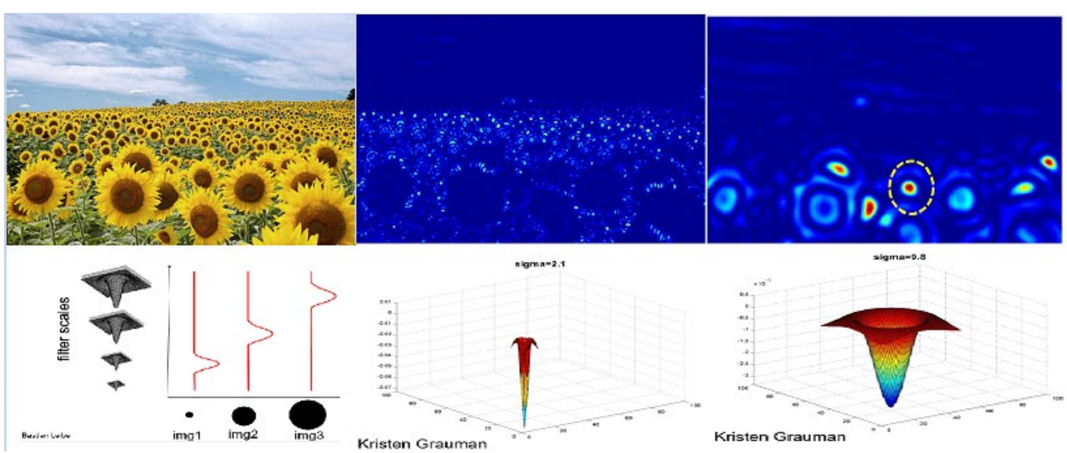


FIGURE 8: LAPLACIAN: A) ORIGINAL IMAGE B) SIMGA = 2.1 C) SIGMA = 9.8 [6]

It is observed that for an image, if sigma, the standard deviation, is smaller, then smaller circles will be selected from the image and if sigma is bigger, then bigger circles will be selected.

Feature Descriptors

The next step in image registration is the feature description part, where the keypoints that have been previously chosen are represented in some way. Some examples of basic features may be [9]:

- Array of RGB values of pixels
- Array of HSL values of pixels
- Histogram of RGB values in keypoint
- Histogram of oriented gradients (HOG)
- Dominant Color Descriptor
- Edge Histogram Descriptor
- Binary Mask
- Combinations of them

What this step does is to make sure that from the most distinctive part of the image the most distinctive way to represent that part is selected. So, if a picture has a wide color range, an HSL feature should be used, if the picture has multiple textures, then an HOG feature should be used in order to differentiate between them. Until a few years ago (2010 - 2014), before neural nets surpassing the State-of-the-Art methods in a lot of fields, the best methods for this step of feature selecting and for the previous step of keypoints detection, were Scale-invariant feature transform (SIFT) and its updated version Speeded up robust features (SURF). These 2 algorithms both make use of edge detection and histograms in order to determine feature points.

Feature matching

This step of image registration consists of matching the features found in the reference image with the features found in the test image. This step is crucial in making the registration, because this is where the correlation between the position of the observed object is done with respect to the other picture. Most methods take advantage of some kind of distance between the feature's values of the keypoints, like the Euclidean Distance [9], but more advanced use, beside those values, the values of the neighbors of the features for a more accurate correlation. In Fig. 9, the matches that were found can be observed and it is noticeable that some of them are miscorrelated.

FIGURE 9: FEATURES BEFORE ELIMINATION [6]

In order to prevent this phenomenon, after the matching, one should keep only the first *N* matches with the lowest distance between them. After this trimming part the correlation will look like the one in Fig. 10.

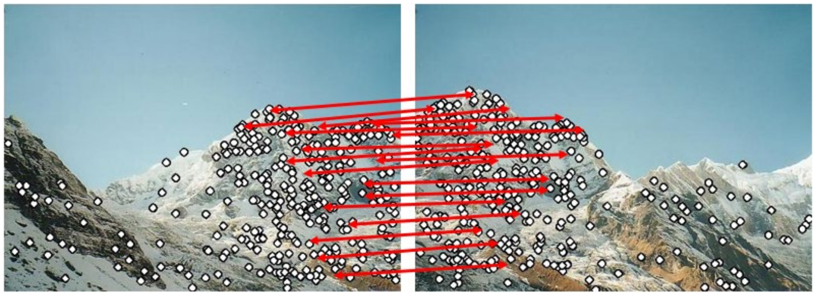


FIGURE 10: FEATURES AFTER ELIMINATION [6]

Now, the selected features have a higher accuracy so, the image will be much better aligned.

Image warping

The final step is to warp the test image in order to transform it to the original one. The warping can be done in many different ways and from a mathematical point of view represents a geometrical transformation. The methods are various and each of them is helpful for different kind of problems.

1. Basic 2D transformations [6]

This method requires a transformation matrix of 3x3 which represents the following operations:

- Translation
- Scale
- Rotate
- Shear

All these transformations are called affine transformations and all have the property of keeping parallel lines parallel [5].

In order to solve this problem, one should apply Least Squares algorithm in order to minimize the equations:

$$x_i' = M_1 * x_i + t \text{ and } y_i' = M_2 * y_i + t$$

where, x_i and y_i represents the coordinates of feature i, M is the transformation matrix and t are the bias.

To solve this equation, 6 parameters are needed, so with 3 pixels, it can be solved [5].

$$\begin{pmatrix} & \cdots & & & \cdots & \\ x_i & y_i & 0 & 0 & 1 & 0 \\ 0 & 0 & x_i & y_i & 0 & 1 \\ & \cdots & & & \cdots & \end{pmatrix} \begin{pmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \\ t_1 \\ t_2 \end{pmatrix} = \begin{pmatrix} \cdots \\ x_i' \\ y_i' \\ \cdots \end{pmatrix}$$

2. Homography [6]

Homography is a plane projective transformation. One typical example of this transformation is viewing one surface from two different points. From the geometrical point of view, this transformation is similar to the previous one, but now 9 parameters are unknown, but just 8 of them are independent, so 4 pixels are required to compute it. This transformation is a distortion of the entire image. Nowadays, libraries that contains the already implemented algorithms for homography exists and can be used easily. For example, RANSAC represents this type of algorithm and can be found in OpenCV module from Python.

3. Local distortions [7]

The most complicated way to transform an image is with local transformations. These transformations are applied to only small patches of the image and are harder to compute. This kind of transformations represents a very big percent of the nowadays research. Usually, they are computed using a gradient field for the transformation map, like in the following image:

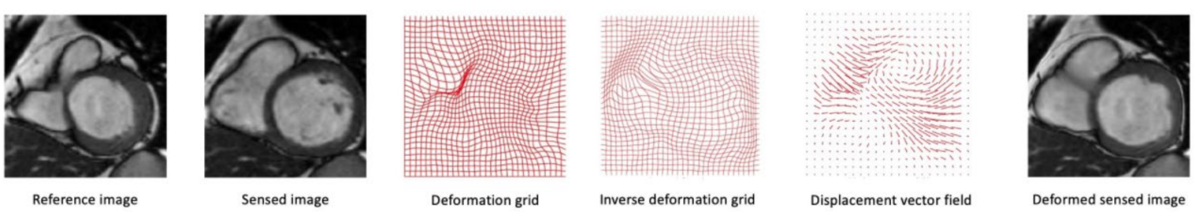


FIGURE 11: LOCAL DISTORTIONS [7]

Hyperspectral Images

Hyperspectral images are images that have been taken at different multiple wavelengths. This procedure is made in order to catch the behavior of the observed objected at single wavelength at a time, instead of multiple wavelengths at a time, like in the RGB images [10].

General description

The visible color spectrum for humans consists in wavelengths varying from 390nm to 700nm [10].

See Fig. 2:

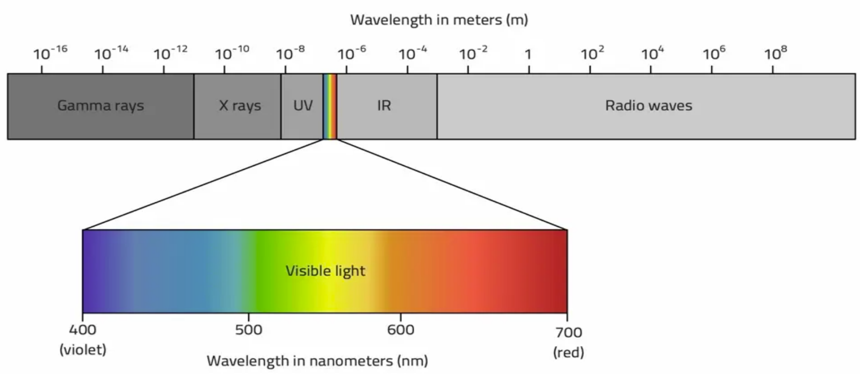


FIGURE 12: WAVELENGTH SPECTRUM [3]

The RGB spectrum takes from the spectrum at once just the wavelengths for red (564–580 nm), green (534–545 nm) and blue (420–440 nm) and what we see is a combination of these three wavelengths [2]. However, the hyperspectral imaging takes every frequency in visible range and plots an image for that object observed just at that frequency. So, if the output of RGB is a linear combination of the three wavelengths, the hyperspectral output is a set of images representing the intensity of every wavelength present in the object.

One common mistake that people make is to mistake hyperspectral images to multispectral images. If the RGB images are on one side and the hyperspectral are on the other, the multispectral images are in between and represent a set of images, but for a fewer frequency. In practice, it is considered that if you have more than 3 wavelengths and less than 40 – 50 is multimodal and if you have more than 50 is hyperspectral [11].

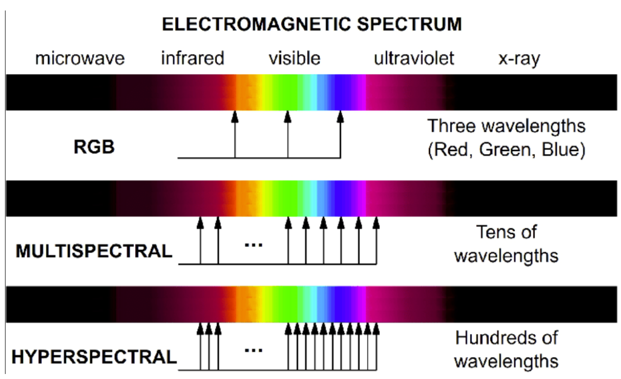


FIGURE 13: A) RGB B) MULTISPECTRAL C) HYPERSPECTRAL [11]

Real life problem

The prior given example regarding the observed algae being extremely hard to see is a real-life problem that is currently trying to be solved by the people at Norwegian University of Science and Technology. They are trying to align the hyperspectral images captured by a special airplane in order to observe the behavior of algae around the coast of North Sea.

The main problem is that the airplane is flying inside our atmosphere. So, in order for light to get from the algae to the lens of the camera mounted on the airplane, it has to go through vegetation, mud, water and air. These components, each affects light in a different way, so the hyperspectral images will take advantage of the frequencies that benefit algae and not the other elements.

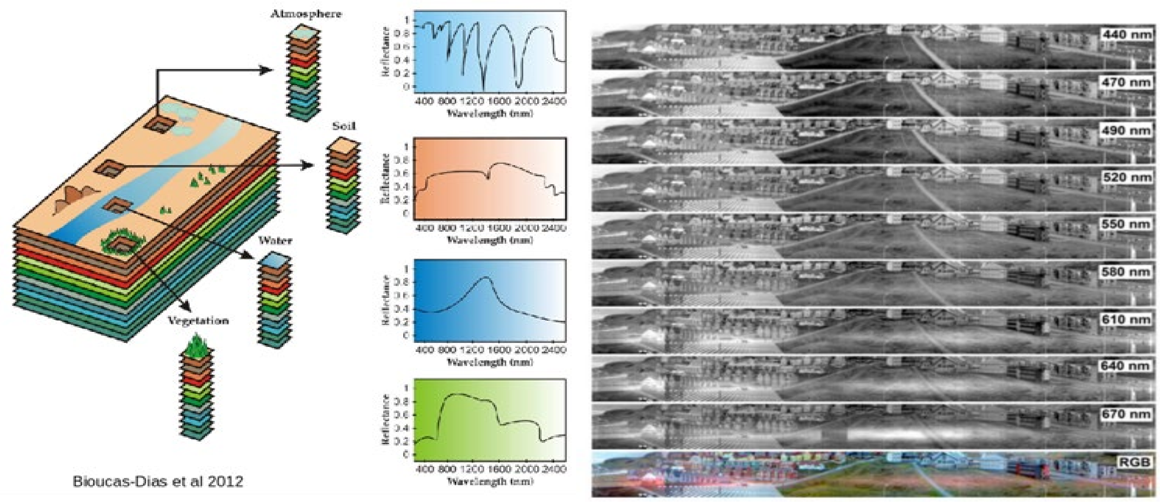


FIGURE 14: LIGHT THROUGH DIFFERENT MATERIALS AT DIFFERENT WAVELENGTHS*

All the images taken from the same area, have to be perfectly aligned in order to match the values for the same pixel. The technological constraints at this moment, compels us to take one single line of pixels at a time when observing the ground. So, before applying the registration, a stitching has to be done between every line of pixels taken. This problem is generated by the way the camera is able to take pictures. The pictures are taken with a specialized camera that has the advantage of capturing all the wavelengths images at a time, but with the cost of observing the ground line by line. You can see the model in the figure 5. Having these limitations, introduces new problems, like dealing with the angles of yaw, roll and pitch of the airplane, so keeping the pixel lines straight is difficult. However, this task is different from image registration, so it does not make the subject of this paper.

**F. Sigernes, M. Surjäsuo, R. Storvold, J. Fortuna, M. E. Grøtte, T. A. Johansen, "Do it yourself hyperspectral imager for handheld to airborne operations", Optics Express, 2018*

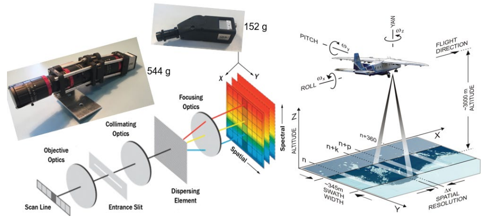


FIGURE 15: NTNU CAMERA SYSTEM*

Hyperspectral Sensors

Hyperspectral sensors have the purpose of collecting information from the environment and transfer it to the computational unit just like any other sensor.[13] The difference between this kind of sensors and the common sensors for capturing images is that these hyperspectral sensors are built for computing a block of images instead of just one image. This block is called hyperspectral data cube and contains the data from all the frequencies used for imaging. However, the frequency is a continuous variable, so that would mean that one dimension of the hyperspectral data cube would be infinite, whatever the size of the chosen spectrum range. To avoid this issue, information from continuous frequency value is plotted together on the same image. So, in fact, each image contains information taken from a small range of wavelengths. What makes a sensor good is how small they can make the interval of frequencies for a single image. The smaller the interval, the better the sensor. Below it is plotted such a hyperspectral data cube:

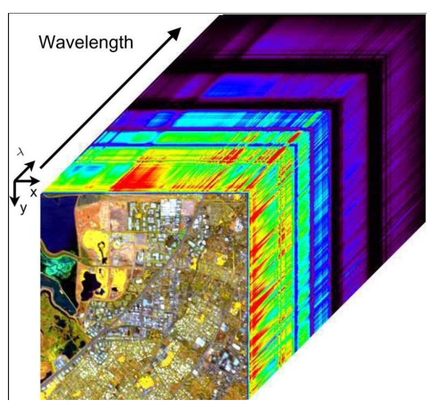


FIGURE 16: HYPERSPECTRAL DATA CUBE**

* *Z. Volent; G. Johnsen; F. Sigernes, "Kelp forest mapping by use of airborne hyperspectral imager", Journal of Applied Remote Sensing. 1. 10.1117/1.2822611. 2007*

***E. Christophe & C. Mailhes & P. Duhamel, "Hyperspectral image compression: Adapting SPIHT and EZW to anisotropic 3-D wavelet coding", IEEE, 2009*

There are different methods in which one can acquire the hyperspectral information. The main 4 methods are [14]:

1. Spatial scanning

This kind of sensors take at one moment one line of pixels, but for the whole spectrum at once. This is also the case of sensors used by the NTNU team in their journey of algae identification. This way of obtaining hyperspectral images is difficult to use because after taking those lines, they need to be put together. This process of putting them back together can introduce errors in the final images because the stitching was not done perfectly.

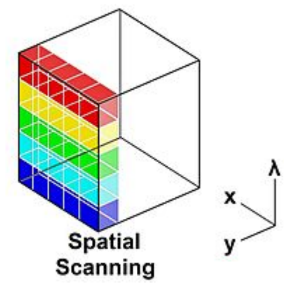


FIGURE 17: SPATIAL SCANNING [14]

2. Spectral scanning

The spectral scanning assumes taking the hyperspectral images one wavelength at a time. One wavelength is in fact a small interval a wavelength, but for the sake of simplicity it is referred as one single wavelength. The upside of using this technique is that you have full control on the frequency that you want to analyze. The downside is that the frequency is given by the chosen lens of the sensor. So multiple frequencies mean having and interchanging the lens one by one. Also, interching the lens means time passing, so this fact makes this method very unsuitable for applications which require taking the images of the frequencies immediately one after another.

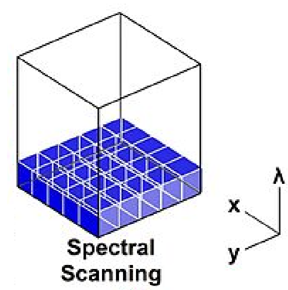


FIGURE 18: SPECTRAL SCANNING [14]

3. Non-scanning

The Non-scanning method represents the best method that it exists. The hyperspectral data cube is computed all at the same time. That means that all the lines and columns for all the wavelengths in the desired interval are computed all at once. The major problem with these sensors and the main reason that they are not use is their price. Being so powerful, they are very expensive. Another disadvantage is that they do not hold any spectral info, but they hold chemical data. In other words, the hyperspectral characteristics are retrieved from the environment through chemical images, not through electromagnetic ones.

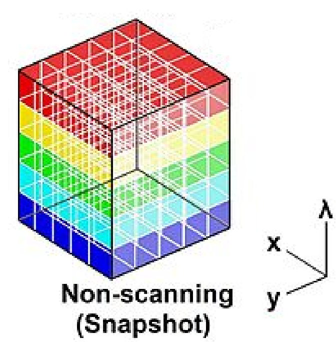


FIGURE 19: NON-SCANNING [14]

4. Spatiospectral scanning

In this type of scanning, the sensor take at one moment of time on 2D image, just like spectral scanning, but the wavelength for every row is different from the others. This is called “rainbow-colored” for this particular reason and it is used more rarely.

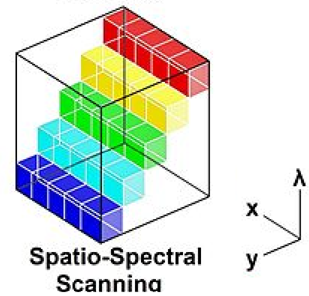


FIGURE 20: SPATIOSPECTRAL SCANNING [14]

Medical use of registration

The use of registration in the medical world comes when talking about scans. Nowadays, there are a lot of medical procedures that requires some kind of body part scans in order to determine the condition that a patient has. For example, some of those procedures are:

- Computer tomograph scans (CT)
- Magnetic resonance imaging scans (MRI)
- Echography
- X-ray

What all those procedures have in common is that they all are part of radiology. So, this pre-processing step of registration is highly used in radiology tasks, because they all involve taking pictures of the patient multiple times.

In this research, the focus of the medical usage is on computer tomograph scans. The reason of choosing this particular topic is the fact that CTs are the most used procedures from the ones that were listed above, so particularizing the problem for them would be the most helpful.

CT procedure

The CT procedure consists in multiple X-rays taken by a machine that is rotating around the patient. The patient is laying on a table and the machine takes several cross-section pictures of the patient. The CT can observe multiple types of tissues and conditions / problems [15]:

- Broken bones
- Blood clots
- Abnormalities of lungs
- Inflammation in the abdomen
- Kidneys stones
- Blocked bowel passage or twisted bowel
- Cancers in different body parts

- Usually, the radiation area is limited by the inflicted organs containing the condition: head, spine, heart, thigh, etc. The whole procedure takes around 5 to 10 minutes and it involves multiple series of scans [16]:

1. Native scan

The first scan is just the normal scan in which the X-ray is applied to the wanted area. The patient is not prepared in any specific way prior to this scan. Usually, this represents the ground truth of the series of images taken by the CT.

2. Arterial scan

For the second scan, a special contrast dye is used by the doctor, in order to highlight the arteries in the patient's body. This dye is given to the patient either by drinking, injection or enema and after a few minutes, another series of images are taken by the machine. The timing of this series is calculated for the contrast substance to be in the patient's arterial system.

3. Veins scan

For the third scan, the patient needs to wait a few more minutes for the substance to travel all the way to the veins system, then the machine will take the third round of photos, highlighting the venous system of the patient.

4. Post contrast scan

After another few minutes, a fourth round of images are taken. Now, the contrast substance had leaved the veins and the machine will highlight now the areas in which the contrast substance is still present.

FIGURE 21: CT SCANS: A) NATIVE B) ARTERIAL C) VENOUS D) POST-CONTRAST *

This whole procedure has the role of capturing how different tissue types are reacting to the X-ray. In this way, the doctor can say whether or not a tissue is cancerous or not. Usually, one series

** L. C. Andrade, "ARP Case Report Nº 14: Primary Hepatic Lymphoma", 2018* of images has around 100-200 photos depending on the length of the observed area of the patient. Taking 4 rounds of photos, this results in a 400-800 pictures per CT scan per patient.

The first step in aligning all the images is to align the ith image of the first series with the other ith images of the second, third and fourth series. This alignment is made on the frontal axis and the saggital axis. After all the images have been aligned between the series, the second step is to align them by the longitudinal axis in order to recreate the 3D model of the analyzed area. This process is crucial for the doctor, because based on the results of the CT, the radiologist will give indications to the surgeons regarding the localization of the tumor of the patient. So, a miscalculated localization means extracting a healthy tissue from the patient or even worse, not extracting the whole cancerous tissue. This can result in the cancerous cells replicating themselves and the effect is that after a few months or years the patient will have cancer again.

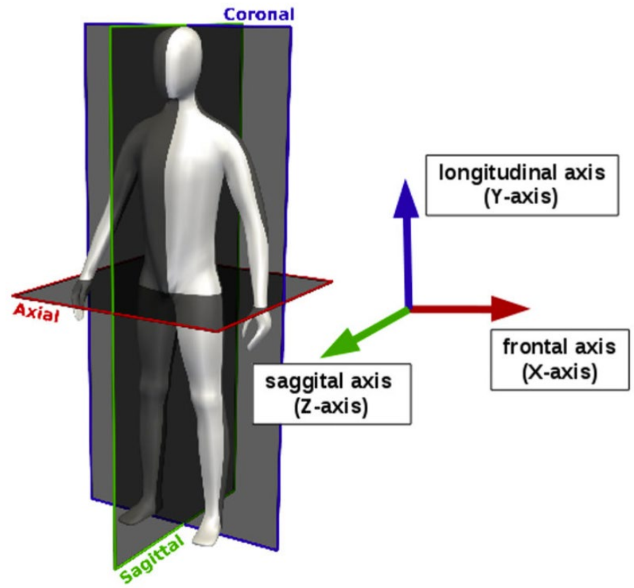


FIGURE 22: CT SCAN AXES *

The current state of the CT scans accuracy, is that every pixel presented in an image represents 1mm to 10mm of cells. This result is not very conclusive and this is where the image registration comes in. If all the images are perfectly aligned, with the help of the other information that the CT is giving, like the Hounsfield units, the doctors may detect lower structures than 1mm from the photos. The Hounsfield unit is a dimensionless unit that measure how much of the energy that was emitted by the machine was absorbed by the cell, so by observing these values, the doctor can say about a cell whether is a bone, a muscle, a tumor etc. [15].

The CT scan emits X-rays that pass through our body. Some of the energy is absorbed, while some of it passes completely through the body and hits the receiver on the other side of the tube. This receiver creates an analogical 3D image based on how much radiations we did not absorbed. This analogical 3D image needs to be transferred to a computer. This is where the voxels come in. A voxel is the smallest unit in the 3D image, similar of how the pixel is to an 2D image. Then, after the 3D image is loaded, several slices are cut from the image cube and the voxels are transformed

**E. Heim, "Large-scale medical image annotation with quality-controlled crowdsourcing", URN: urn:nbn:de:bsz:16-heidok-246418 , 2018* to pixels. If the registration is along the longitudinal axis, then the number of slices in the cube is similar to the number of images for a single wavelength in a hyperspectral data cube.

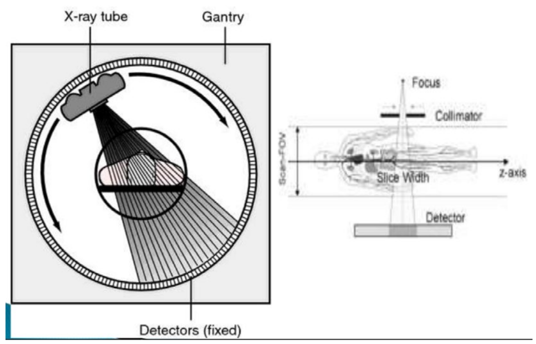


FIGURE 23: CT PRINCIPLE [15]

Hyperspectral and State-of-the-Art

Although the CT scans appear to have no connection with the hyperspectral images taken from the airplane, both approaches profit from the same principle: observe different behaviors of the object when stimulated with different inputs. In the case of the algae, the inputs are the wavelengths of the light that hit the lens of the camera and on CT the inputs are the levels of contrast substance in the patient's body.

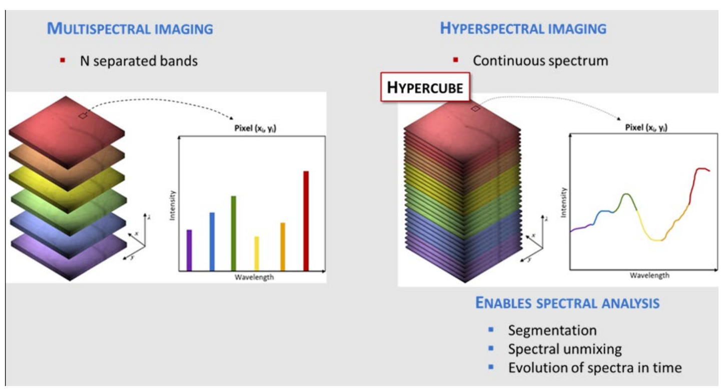


FIGURE 24: MULTISPECTRAL VS. HYPERSPECTRAL*

CT images represent some "multispectral" data, for which the wavelength of the light is replaced by the concentration of contrast substance in cells. For algae, taking more than 50 images at different wavelength, the Hypercube can be considered to be continuous, so the value of pixel (i,j) can be considered to be a continuous variable by approximation. Solving the problem of image registration on medical data, where accuracy must be extremely high, solves automatically the problem of registration on geography mapping, where the accuracy can be lower. Given the fact that by now, the people at NTNU have not solved yet the issue of image stitching, one way to work for this kind of problems is to work on medical images.

**S. Feliz, "Ct scan", https://www.slideshare.net/shreyacathe/ct-scan-62017319*

After searching for State-of-the Art algorithms for implementation image registration, I drew the following conclusion: there are several State-of-the-Art algorithms and they differ from one another by the approach of the problem.

Classical feature-based Methods

The classical methods are represented my any algorithm that it does not involve machine learning or neural networks. These methods follow the standard steps of image registration that were described above. The main particularity of them are the transformation matrix and their geometrical aspect of computing them. Two of those methods are SIFT and SURF combined with a typical homography to distort the image as a whole. The keypoints are selected while looking for edges and corners and the features are based on histograms. SURF was considered to be a State-of-the-Art method for registration for a long time, until machine learning took first place. However, SURF and SIRF can still be found in Python libraries in older versions. In the newer versions they are not free anymore, but old versions of Python still have them as free. [6]

Regarding the feature matching part of the algorithm, classical approaches may include some simple brute-force matching which compute the distance between every feature in the reference image and the test image and then select just the best batches, the matches with the lowest distance. The distance metric is given by the user and it can be Euclidean, Manhattan or any other distance. Another method is to take into account the neighbors of the features, resulting on a K-nearest neighbors (Knn) matching which compares the neighbors to in order to identify the best matches. After all matches are computed one have to set a threshold in order to retain just the best *N* matches, not all of them. BFMatcher and KNNMathcer are functions from OpenCV which implements those methods. [7]

The warping of the image is done by a homography like the ones discussed earlier. RANSAC algorithm can be used in order to determine and eliminate outliers and them the homography matrix can be formed using Least Squares. Again, OpenCV library from Python has all these functions implemented.

Machine learning methods

With all the hype of machine learning surpassing every classical method, people found ways to improve the classical methods of registration too. They found different ways to introduce machine learning in multiple places in the registration routine.

The first thing that they did, was to replace the whole part of registration with a network. In 2016, D. DeTone[20] replaced the whole algorithm with a neural network similar to VGG networks and learned the whole homography in this network. The output of the network were the 8 unknown values for computing the homography and nothing else. The network was supervised, so labeled data was needed. The network had the following layers:

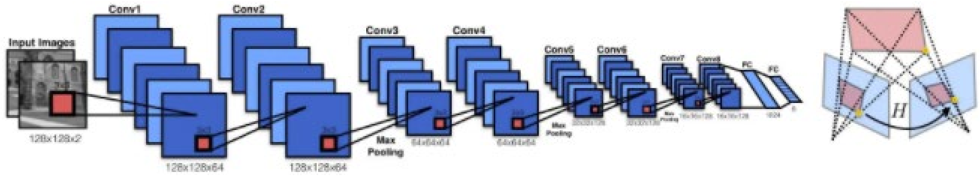


FIGURE 25: REGRESSION HOMOGRAPHY [20]

Computing the homography matrix and needing labeled data, results in a prior computed homography matrix for each pair of images, but after that was achieved the network could be trained. Here is the logic behind the model:

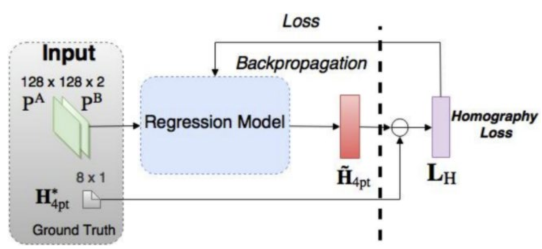


FIGURE 26: SUPERVISED LOGIC [20]

Others approaches involved reinforcement learning: R. Liao with a paper in 2016 performed the implementation of a network that was replacing the whole registration as well [19].

The drawback of all these methods is the way of transforming the image: rigid transformation. The images are transformed as a whole and from a medical point of view this thing is not wanted at all because tumors can change their shape independent from the shape of the other organs. So, this is the reason for which I did not choose any one of them in order to improve.

Another kind of methods are represented by complex transformations, or non-rigid transformations. These methods work pretty well on medical images and for example Julian Krebs when published his paper in 2017 regarding a neural network with reinforced learning, tested his network on a dataset of prostate MRI images. So, these transformations are perfect for the purpose of this paper. However, when speaking with doctors and people from the medical field, I have discovered the following problem: doctors does not trust any method which involves generating the registered image by a network. They have said that in any circumstance they would trust some picture generated by artificial intelligence. So, even though this is not a mathematical problem, but an ethical one, as long as the doctors don't want to apply them in real life, they have no real purpose.

After searching for a method that took the best of both worlds, I have found the following paper which uses machine learning in order to find the transformations, but without generating the registered image using machine learning.

Best method for this project

The paper is called "Multi-Temporal Remote Sensing Image Registration Using Deep Convolutional Features" and it was written by Z. Yang , T. Dan and Y. Yang in 2018. After reading it and understanding it, I was able to find some possible improvements in particularizing the problem for the medical field. However, this method proves to be efficient in the other field too, so giving geographical and medical data for this method appears to work rather well. All the information from this chapter is based on the mentioned paper. [17]

Feature selection

The main particularity of this paper is the method used for finding the features of images. Instead of using SIFT or other method, the authors take the features from the pooling layers of a VGG16 network trained on ImageNet.

The reason for taking the output of those layers is that because of the shape of the maps, the features will represent different patches with different sizes from the image. For example, the first pooling layers will produce a feature for a 8x8 patch, the next pooling layer will produce features for an 16x16 patch and so on. Even though the net was trained for a classification, the authors use it in other way. The network was trained on ImageNet, so general images are being used for trained and the dataset that the authors use contains images captured by a drone from 50-100 meters altitude showing land and water areas, so the training of the network is in relation with the dataset used. Both, the reference image and the test image are given as input to the network

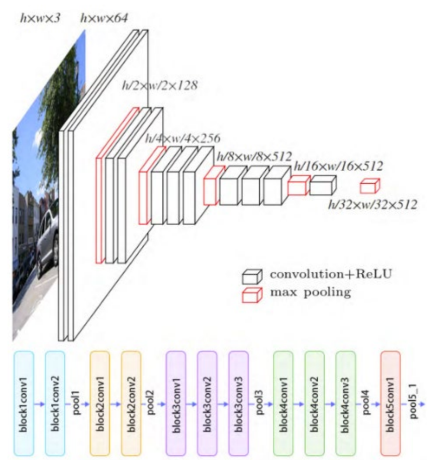


FIGURE 27: VGG16 STRUCTURE [17]

at the same time and the features are selected at the same time. The input images have the size of 224x224 pixels and a feature for a patch has the size of 7x7x512 for a patch of 32x32 patches. 512 is the number of maps that are produced by the neural network. As you can see from the Fig. 21, not all features have the number of maps equal to 512. Some of them have the size of 256. These features must be rescaled in order to match the size of the other ones. Three pooling layers will be considered for taking the features. The distance from one point to another will be the sum of the Euclidean distances from one feature to another. Below is a figure in which each dot represents a feature. Their color determines the pooling layer from which they got extracted from. The green dots were extracted from the first pooling layer and represents the feature of a 8x8 patch, the blue ones are from the second pooling layer and represents a 16x16 patch and the prink one is extracted from the last pooling layer and represents a 32x32 patch. The reason behind selecting features from multiple pooling layers is that selecting them this way, the algorithm will have feature for different sizes entities from the images.

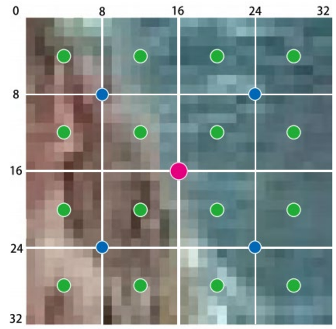


FIGURE 28: FEATURES DISTRIBUTION

Feature matching and transformation

For function that represents the matching is a combination of two costs: [17]

- one cost represents the convolutional feature cost matrix and cares about the Euclidean distance from a point to another:

$$C_\theta^{\text{conv}}[m,n] = \begin{cases} \dfrac{d(y_m, x_n)}{d_\theta^{\max}}, & \text{condition 1} \\ 1, & \text{otherwise.} \end{cases}$$

- the other one is the geometrical cost and cares about the angles and the distance between the points surrounding the current features and the feature:

$$C^{\text{geo}}[m,n] = \frac{1}{2} \sum_{b=1}^{B} \frac{[h_m^y(b) - h_n^x(b)]^2}{h_m^y(b) + h_n^x(b)},$$

These 2 costs form a loss function that needs to be minimized in order to find the best pairs. In order to find the best pairs a Linear Assignment Problem solver is used. It is the same problem as: assign some agents to some tasks in order to minimize the total time of execution. This step comes with a prior probability matrix which is:

$$P_R[m,n] = \begin{cases} 1, & \text{if } y_m \text{ and } x_n \text{ are corresponding} \\ \dfrac{1-\epsilon}{N}, & \text{otherwise.} \end{cases}$$

Having the minimal cost of all pairs of pixels from both images: X_i and Y_i, we now must transform the test image in order to look like the reference image. This procedure is based on equation:

$$Z = Y + GW$$

Where, Z is the registered image, Y is the test image, G is a matrix generated by a Gaussian radial basis function and W are the parameters that need to be learned. The learning of these parameters is done in an Estimation-Maximization manner, in which the probability of choosing correctly

$$P_O[m,n] = P^{\text{old}}(m|x_n) = \frac{P_R[m,n] g_m(x_n)}{p(x_n)}$$

the pairs in the Jonker-Volgenant assignment algorithm from a Gaussian Mixture Model is estimated and the maximization of the likelihood function is computed.

E-step: computing probability matrix P_o:

Where, g_m is a GMM probability:

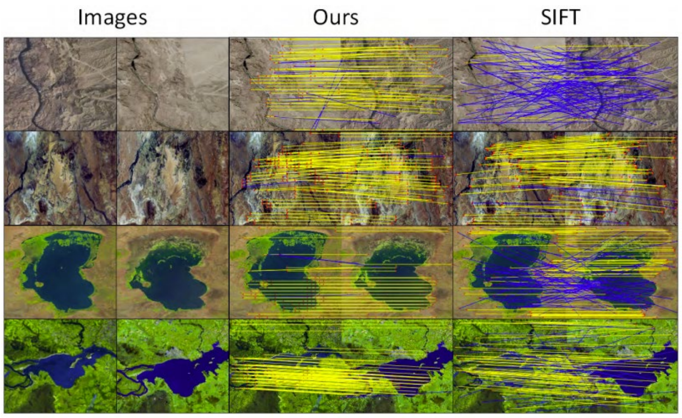
$$g_m(x) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{1}{2\sigma^2} \|x - y_m\|^2\right)$$

The M-step of this procedure updates the parameters of those functions such as σ^2 (the variance of GMM model) and W

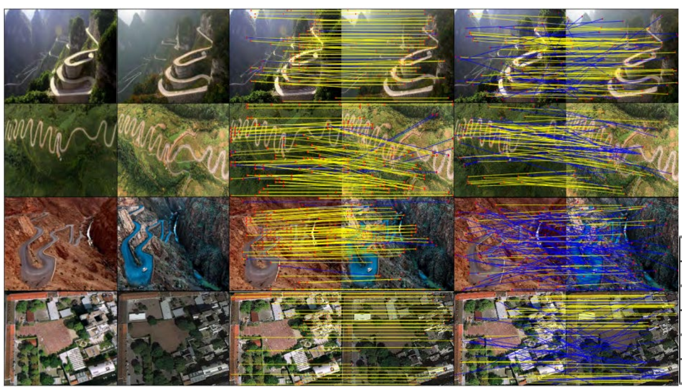
This procedure results in slowly increasing the similarity between the reference image and the registered image.

Results

The authors claim that their algorithm performs better than SIFT because they solve on of SIFTs biggest problems: the incorrect matching of many features. While using SIFT, one individual may remain with too few correctly identified features, while 90% of them are wrong. Below you can see with yellow lines the correctly identified features and with blue lines the misidentified ones:



dataset	index	Ours	SIFT
Satellite	avg.	95.65	71.71
	min.	88.24	30.21
	max.	100.00	93.48
	std.	3.01	19.03
UAV	avg.	93.37	42.94
	min.	86.27	12.90
	max.	100.00	78.50
	std.	4.166	19.12



Method	RMSD	MAD	MED	STD
Ours	12.63	13.20	6.78	5.73
CPD	24.01	30.12	7.51	14.56
GLMDTPS	27.40	35.28	8.26	15.26
GL-CATE	15.22	19.23	6.69	13.77
PR-GLS	23.52	29.61	7.75	11.67

FIGURE 29: PAPER RESULTS VS SIFT [17]

As you can see this method outperforms SIFT and thus making it one of the currently State-of-the-Art methods.

Experience and results

My first experience regarding this project was to determine what I wanted to do and for this task my coordinator professor, I. Necoara, really helped me by introducing me to a research group of people that were studying hyperspectral imaging. This group of people was formed by some professors and PhD students from the NTNU faculty, some professors from University of Bucharest and some doctors from Coltea Hospital. Their task is to find more about hyperspectral imaging and after listening to every step of this pre-processing, being just a master student, I had the opportunity to chose if I want to get involved and where I want to get involve, because any help was well received. Being at the Artificial Intelligence master and already loving image processing, I chose to get involved and that is how I ended up liking image registration. After I have found out that the courses from the master were really similar to this topic, I decided to make this my research project and to get involved much more.

One of the constant things that I did was to meet once every two weeks with the other people in order to exchange information. That is how I learned more about the need of this image registration. The people from NTNU need this in order to observe the algae from the North Sea, but this step will come in play later because now they are at the point where they need to stich the lines of pixels gotten from the airplane. Talking with them, I have found out that multiple operations have to be done on hyperspectral imaging in order to use them. Such operations are: Stitching, because they use a spatial sensor for obtaining the images, registration because the plane from which the images are taken is vibrating and moving, super resolution, because the plane is very high and the algae is on the bottom of the ocean, so they cannot see them properly, etc. Coltea Hospital need this step of registration right now, because in their process of enhancement of resolution of CT images, their blocking point is represented by the registration. So, I turned my interest for medical imaging.

After speaking with a radiologist doctor and with several students from the hospital, I learned more about how the CT are done, about the contrast substance, the radiation that the machine is emitting, the Hounsfield units, the transformation from voxels to pixels and so on. Visiting Coltea Hospital, the radiologist doctor showed me series of images from the CT and decided to share with me the CT scan images of 100 anonymous people, which means 100 people * 150 images per series * 4 series = 60,000 CT images to work on.

Another aspect of my research was understanding and implementing the code for the chosen method of image registration. After many searches, I managed to find a python code which was not working, but was implementing the method, so I could not run it until I understood it in order to modify it. The code was written in Python, so considered the laboratories from the courses of this semester, I was very fortunate to find it. After I modified it, I have tested it on some examples of land and water images extracted from the Computer Vision course and the results were great. The registration is working perfectly on general images. Here is an example on a frame:

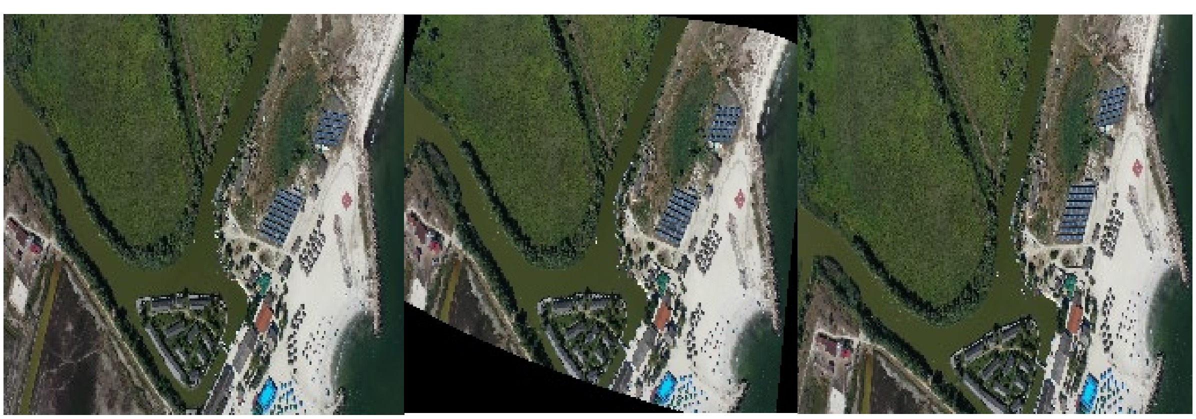


FIGURE 30: A) REFERENCE IMAGE B) REGISTERED IMAGE C) TEST IMAGE[6]

The image on the left is the “ground truth” and the image on the right is the test image which has to be transformed. The center image is the registered image, which means the right image warped to fit the left image. The black areas appear because the felt corner from the reference image is missing in the right one, so it was filled with black. In order to see how good the registration was the code had a feature which produced besides the registered image, one “checkered” image which contained alternating patches from both the original and the registered images. Here is the result:

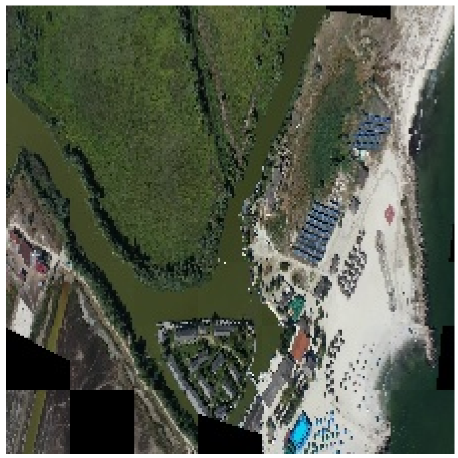


FIGURE 31: CHECKERED IMAGE

The images are almost perfectly aligned, especially in the middle part. Even though they can be seen, even in the middle of the image there are squares belonging one tot the reference, one to the registration.

Until now, I have not received the dataset of medical images from Coltea Hospital, but I wanted to test the algorithm on medical images too. The results were not great at all because the VGG16 that was being used is trained on ImageNet, which is a general image database, not a medical image one.

Conclusions and future work

Work until now

During this semester I had accommodated myself with the subject of hyperspectral images and with registration, I managed to track down most of the State-of-the-Art implementations for registration, I understood the need of registration in real life applications and I managed to take part in a project involving this pre-processing step along other passionate people. After implementing the algorithm and observing a real dataset of CT scans, I thought of some improvements on the actual algorithm that may benefit the project.

1. The network from which the features are extracted is a VGG16 trained on ImageNet. A particularization of the training dataset and also of the network structure, should increase the accuracy of registration on medical images.
2. I should consider specific areas of the CT scans from which the features are selected: for examples it is more likely for bones like the ribs or the spine to maintain their shape throughout the whole series than it is for the lungs which expand when someone breath
3. Modify hyper-parameters from the algorithm after changing the network.

Future work

The goal for the second semester is to test the algorithm on the dataset gotten from Coltea Hospital and to modify it in relation with the dataset. Also, in the second semester I want to obtain a dataset from the NTNU team with the algae and to test the algorithm there as well. If all the experiments go well, more investigations about the improvement of this algorithm should be made. These improvements may come either from the computer vision point of view, machine learning point of view, or even medical point of view.

Conclusions

The work until now has proven to be an amazing work because I had the opportunity to work with amazing people and to learn such wonderful stuff. The fact that my work could help a lot of people one day, keeps me motivated to continue with this research project. I am glad to say I have accommodated myself with this topic and this the State-of-the-Art methods of this subject, so based on all the things that I have described above, I consider that my objective for this semester was achieved with great success.

The State-of-the-Art algorithm that I have found, seems suitable for my project and for now I am more than happy with its results. However, in the future semester, more tweaking and experiments have to be done for it to prove its efficiency.

Bibliography

1. F. Oliveira & J. Tavares, *"Medical image registration: A review."* Computer methods in biomechanics and biomedical engineering, 17. 73-93. 10.1080/10255842.2012.670855., 2014

2. O. Mendoza-Schrock, J. A. Patrick, E.P Blasch, *"Video Image Registration Evaluation for a Layered Sensing Environment"*, Proceedings of the IEEE 2009 National Aerospace & Electronics Conference (NAECON), Dayton, OH, USA, 2009, pp. 223-230, doi: 10.1109/NAECON.2009.5426624., 21-23 July 2009

3. M. H. Schmidt, *"Frameless stereotactic image-guided C1-C2 transarticular screw fixation for atlantoaxial instability: review of 20 patients"*, Journal of Spinal Disorders & Techniques, doi: 10.1097/01.bsd.0000169443.44202.67, 2005

4. L. G. Brown, *"A survey of image registration techniques (abstract)"*, ACM Computing Surveys archive, volume 24, issue 4, pages 325 – 376, December 1992

5. B. Zitova, J. Flusser, *"Image registration methods: a survey. Image and Vision Computing",* Image and Vision Computing, Volume 21, Issue 11, October 2003, Pages 977-1000, 2003

6. M. Leordeanu, *Computer Vision slides, Course 8: Alignment and matching*

7. E. Kamoun, J. Joslove, "*Image Registration: From SIFT to Deep Learning*", https://www.sicara.ai/blog/2019-07-16-image-registration-deep-learning, 2019

8. S. Panigrahi, *"What are Hyperspectral Images?"*, https://towardsdatascience.com/what-are-hyper-spectral-images-a5de5d9fa91, 2019.

9. B.S. Manjunath (Editor), Philippe Salembier (Editor), and Thomas Sikora (Editor): *"Introduction to MPEG-7: Multimedia Content Description Interface",* Wiley & Sons, April 2002 - ISBN 0-471-48678-7

10. C.I Chang, *"Hyperspectral Imaging: Techniques for Spectral Detection and Classification."* Springer Science & Business Media. ISBN 978-0-306-47483-5., 31 July 2003

11. H. Grahn; P. Geladi, *"Techniques and Applications of Hyperspectral Image Analysis. ",* John Wiley & Sons. ISBN 978-0-470-01087-7., 27 September 2007

12. J. Zabalza, *“Feature Extraction and Data Reduction for Hyperspectral Remote Sensing Earth Observation”*, 2015

13. L. Guolan, F. Baowei (2014), *"SPIE – Journal of Biomedical Optics – Medical hyperspectral imaging: a review"*,Journal of Biomedical Optics. 19 (1): 010901. doi:10.1117/1.JBO.19.1.010901. PMC 3895860. PMID 24441941, 2014

14. N. Hagen, M. W. Kudenov, *"Review of snapshot spectral imaging technologies"*, Optical Engineering. 52 (9): 090901. Bibcode:2013OptEn..52i0901H. doi:10.1117/1.OE.52.9.090901. S2CID 215807781., 2013

15. R. Sreedhar, “Basics of CT Scan”, https://www.slideshare.net/sreedharrao313/microsoft-power-point-ct-scan, 2012

16. S. Chiewvit, T. Jiranantanakorn, *“Detection of Recurrent Colorectal Cancer by 18F-FDG PET/CT Comparison with Contrast Enhanced CT Scan”,* J Med Assoc Thai Vol. 96 No. 6, 2013

17. Z. Yang, T. Dan, Y. Yang , *“Multi-Temporal Remote Sensing Image Registration Using Deep Convolutional Features”*, I*EEE Access*, vol. 6, pp. 38544-38555, 2018, doi: 10.1109/ACCESS.2018.2853100., 2018

18. D. DeTone, T. Malisiewicz, A. Rabinovich, “*Deep Image Homography Estimation*”, arXiv:1606.03798, 2016

19. R. Liao, S. Miao, P. de Torunemire, *“An Artificial Agent for Robust Image Registration”*, arXiv:1611.10336, 2016