Supervised and unsupervised image registration

Adama BARRY, Mathieu GRASLAND

January 12, 2021

Abstract

This paper aims to review image registration and 3 methods of image registration. Image registration is the process of overlaying two or more images of the same scene taken at different times, from different viewpoints, and/or by different sensors. It geometrically aligns images; sensed images using the reference image. We present the functioning of the 3 algorithms which are the Coherent Point Drift, Mutual Information and the ORB. Our main goal is to choose the best one to use for a plant dataset given by *INRAE*, thus we will compare them according to there performances for registration but also in computing time.

Key words : *Image Registration, correspondence, matching, alignment, rigid, non-rigid, point sets, Coherent Point Drift, Gaussian mixture models, coherence, regularization, EM algorithm, Feature detection/matching, mapping function, cross-correlation, mutual information, keypoints, corners, FAST, BRIEF, Harris measure, moments.*

Introduction

In the last twenty years, image acquisition devices have undergone rapid development. In the context of the growing amount and diversity of images, it invokes the research on automatic image processing.

We need image registration as a part of image processing, it is a crucial step that aims to enhance the quality of the possible analysis like image fusion, change detection, and multichannel image restoration.

Image registration, as defined in [1], is the process of overlaying two or more images of the same scene taken at different times, from different viewpoints, and/or by different sensors. It geometrically aligns images; sensed images using the reference image. It implies the estimation of an optimal geometric transformation to bring homologous points of two images as close as possible.

Typically, registration is required in remote sensing (multispectral classification, environmental monitoring, change detection, image mosaicing, weather forecasting, creating super-resolution images), in medicine (monitoring tumour growth, treatment verification, comparison of the patient's data with anatomical atlases), in cartography (map updating), and computer vision (target localization, automatic quality control), to name a few.

This paper aims to compare three different methods of image registration. We will focus on **CPD : Coherent Point Drift**, **Information methods** and, **ORB : Oriented FAST and Rotated BRIEF**. The paper is part of a bigger project on pathogen detection on rapeseed leaves, thus we will use registration on images provided by *INRAE (Institut national de la recherche agronomique)*. It is a French agronomic research organisation.

Our goal is to pick the best registration method in order

to use it to carry on pathogen detection and analysis.

In the first part, we will explain the theory and details of image registration, describe the operating of CPD, Information methods and ORB. We will also compare them. In the second part, we will explain how we evaluate the performances of the previous methods and use this evaluation to choose the most suitable for our situation.

1 Image registration methods

In practice, the algorithms linked to the registration methods are supposed to be efficient enough in terms of computational cost and the transformation should be tractable. High dimension images should be handled as easily as smaller ones. One last requirement should be robustness to noise, outliers, or missing points that might occur in the acquisition step.

These methods can either be rigid or non-rigid. A rigid transformation only uses translation, rotation, and scaling of images. It means that rigid methods are sensitive to the spatial perspective. The non-rigid transformation includes the rigid transformations however they are improved by managing perspective transformations. In real life situation, we will favour the non-rigid case because it can handle deformable motion tracking, shape recognition, and medical image registration. However, the true underlying non-rigid transformation model is often unknown and challenging to model.

Since our image dataset is focused on leaves, we will need non-rigid methods. **ORB and mutual information** are in this category, whereas **CPD** can be seen as rigid and non-rigid depending on how you use it. We can also discriminate the algorithms whether they are area-based or features-based. In the area case, the methods use the distribution of pixel intensities to operate registration, whereas in the features case, it uses rules to detect salient and distinctive object like edges, contours, corners or line intersections. **CPD and mutual information** are area-based whereas **ORB** is feature-based.

Registration methods can be divided into 4 steps in the feature detection based, and 3 steps for area-based ones. Although, in the area-based case, the first step is the merge of the two first steps of the feature detection based methods.

These steps consist of *feature detection*, where the algorithm manually (by hand) or automatically detects features previously enumerated represented by control points (CPs). CPs can be centre of gravity, the end of a line or distinctive points. The second step is *feature matching*. Using the CPs of both images, descriptors and similarity measures are used to make correspondence. The third step is the *estimation of the transformation model*. Mapping function aligns the sensed image with the reference using the correspondences. The last step is the *image re-sampling and transformation*. It uses interpolation to infer missing or outlier pixels due to the transformation.

In area-based methods, *feature detection* and *feature matching* are mostly done simultaneously. These methods are applied when the details of a picture are not prominent and the information is given by pixel rather than shape or structure.

1.1 Coherent Point Drift

Coherent Point Drift is a probabilistic multidimensional point sets registration. It is presented as robust, rigid and non-rigid methods [2]. It is also considered as an area-based method. The algorithm uses two sets of points and considered the problem as a probability density estimation problem. One set represents the Gaussian Mixture Model (GMM). The other one represents the data points. A Gaussian mixture model is a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters.

The first step of this process is to fit the GMM centroids to the data by maximizing the likelihood of the GMM. At the maximum, the points sets are aligned and the algorithm can find the correspondences using the probabilities derived after the update of the parameters of the GMM in a second step. This probabilistic method is common to other algorithms.

The CPD algorithm stands out by forcing GMM centroids to move as a group in order to maintain the topological structure of the point sets. It means that it forces the points set to keep some inherent properties. In the rigid case, the GMM centroid locations are explicitly re-

parametrized using a coherence constraint. In the non-rigid case, the displacement area is regularized.

1.1.1 The Gaussian Mixture Model Fitting

The notations are :

- D - dimension of the point sets,
- N, M - number of points in the sets,
- $\mathbf{X}_{N\times D} = (x_1, ..., x_N)^T$ - the first point set (the data points) which can be seen as the reference image,
- $\mathbf{Y}_{M\times D} = (y_1, ..., y_M)^T$ - the second point set (GMM centroids),
- $\mathcal{T}(\mathbf{Y}, \theta)$ - Transformation $\mathcal{T}$ applied to $\mathbf{Y}$, where θ is a set of the transformation parameters,
- $\mathbf{I}$ - identity matrix
- $d(a)$ - a diagonal matrix formed from the vector a.

The algorithm aim to maximize the likelihood. We need to define the GMM probability density function :

$$p(\mathbf{x}) = \sum_{m=1}^{M+1} P(m)p(\mathbf{x} \mid m) \tag{1}$$

where $p(\mathbf{x} \mid m) = \frac{1}{(2\pi\sigma^2)^{D/2}} \exp^{-\frac{\|\mathbf{x}-\mathbf{y}_m\|^2}{2\sigma^2}}$ the normal density with σ^2 equal for all points. $P(m) = \frac{1}{M}$ which means that all GMM components have equal membership probabilities in the gaussian mixture.

Equivalently, we minimize the negative log-likelihood function derived from the previous equations and assuming independent and identically distributed data:

$$E\left(\theta, \sigma^2\right) = -\sum_{n=1}^{N} \log \sum_{m=1}^{M+1} P(m)p(\mathbf{x} \mid m) \tag{2}$$

It results of the estimation of θ and σ using the Expectation Maximization (EM) algorithm. It uses bayesian analysis and to be more precise the Bayes' theorem, in order to compute *a posteriori* probabilities distributions, the expectation. Then parameters are updated by minimizing the negative log-likelihood updated :

$$Q = -\sum_{n=1}^{N}\sum_{m=1}^{M+1} P^{old}\left(m \mid \mathbf{x}_n\right) \log\left(P^{new}(m)p^{new}\left(\mathbf{x}_n \mid m\right)\right) \tag{3}$$

the step of maximization is completed and the algorithm alternate between these two steps until there is no further maximization of the likelihood.

1.1.2 Correspondances and Registration

The correspondances between $\mathbf{X}$ and $\mathbf{Y}$ are derived from the posterior probabilities. Equation (3) can be rewritten as :

$$Q\left(\theta, \sigma^2\right) = \frac{1}{2\sigma^2}\sum_{n=1}^{N}\sum_{m=1}^{M} P^{\text{old}}\ \left(m \mid \mathbf{x}_n\right) \left\|\mathbf{x}_n - \mathcal{T}\left(\mathbf{y}_m, \theta\right)\right\|^2 + \frac{N_{\mathbf{P}} D}{2} \log \sigma^2 \tag{4}$$

where $N_{\mathbf{P}} = \sum_{n=1}^{N}\sum_{m=1}^{M} P^{\text{old}}\ \left(m \mid \mathbf{x}_n\right)$, and $\mathcal{T}$ is to specify according to rigid or non-rigid registration. The function $\mathcal{T}$ is optimized during the maximization step.

In rigid registration, $\mathcal{T}$ is defined as $\mathcal{T}(y_m; \mathbf{R}, \mathbf{t}, s) = s\mathbf{R}y_m + t$ with $\mathbf{R}_{D\times D}$ a rotation matrix, $t_{D\times 1}$ a translation vector and s a scaling parameter. As a rotation matrix, $\mathbf{R}$ should be an orthogonal matrix i.e. $\mathbf{R}^T\mathbf{R} = I$, $det(\mathbf{R}) = 1$

In non-rigid registration, using the Tikhonov regularization, the transformation $\mathcal{T}$ is defined as $\mathcal{T}(\mathbf{Y}, v) = \mathbf{Y} + v(\mathbf{Y})$. The function v is also optimized during the maximization step, with regularization. In practice, it implies the use of a matrix that works like a low-pass filter. Thus, when there are too many differences of transformation between close points, this matrix allows regularization to make transformation more coherent as a group.

1.2 Information methods

The **Mutual Information (MI) and Cross-Correlation (CC)** methods are area-based methods. The paper of Zitova claims MI is a leading technique in multimodal images [1]. These methods uses correlation-like methods and merge the *detection* step with the *matching* one.

1.2.1 Cross-correlation

The classical way of extracting information from an image is to use cross-correlation (CC) in order to directly

match image intensities, without structural analysis.

$$CC(i,j) = \frac{\sum_W (W - E(W)) \left(I_{(i,j)} - E\left(I_{(i,j)}\right)\right)}{\sqrt{\sum_W (W - E(W))^2} \sqrt{\sum_{I_{(i,j)}} \left(I_{(i,j)} - E\left(I_{(i,j)}\right)\right)^2}} \tag{5}$$

The CC is computed between two pair of windows (W) from the sensed and reference image.

This measure implies that this method is very sensitive to intensity changes brought by noise, different scene times or sensor type. However, registration with the CC measure can exactly align mutually translated images, slight rotation and scaling.

The computational cost grows very fast when the image information increase. In this case, the method can be simplified using the sum of the absolute differences of the image intensity. When it exceeds a threshold to fix, the candidate pair of windows are rejected.

1.2.2 Mutual Information

MI is derivated from the information theory, it measures a statistical dependency between two pixel datasets. When there are different modalities this method is convenient, colour images fall in this category. It uses the theoretical statistical entropy of random variables. Given two random variables X and Y, the formula of MI is :

$$\mathrm{MI}(X,Y) = H(Y) - H(Y \mid X) = H(X) + H(Y) - H(X,Y), \tag{6}$$

where $H(X) = -E_X(\log(P(X)))$ represents the entropy and $P(X)$ the probability distribution of the random variable X. Then MI is maximised using gradient descent optimization. This method works with the entire image data and there is no need to preprocess pixels.

1.2.3 Mapping function and Registration

With both methods, features have been detected and matched simultaneously. Since scale, rotation and more is to be modified for the sensed image, the mapping function should be defined. There are two categories of mapping function. *Global models* uses all the control points (CPs) found in order to estimate one set of parameters for the transformation. *Local models* uses a composition of mapping function where the parameters depend on the image's region.

Global mapping models consists of preserving the shape of the set of points. The similarity transform find the best rotation, translation and scale modifications using 2 CP :

$$\begin{aligned} u &= s(x\cos(\varphi) - y\sin(\varphi)) + t_x \\ v &= s(x\sin(\varphi) + y\cos(\varphi)) + t_y \end{aligned} \tag{7}$$

where s represents the scaling parameter, t_x and t_y the translation on abscissa and ordinate axis, and φ the angle of rotation.

Another transformation is the affine transform which uses 3 CPs and preserves straight-line parallelism.

$$\begin{aligned} u &= a_0 + a_1 x + a_2 y \\ v &= b_0 + b_1 x + b_2 y \end{aligned} \tag{8}$$

In general, more CPs than needed are used to estimate the parameters of the mapping function. There are often computed using means of the least-square minimizing the sum of squared errors for every CPs. This method was proven effective for satellite images.

Local mapping functions are used to transform local deformed images. The previous mapping will not be accurate for the whole set of points. The images are divided into windows consisting of enough matching CPs and the previous methods are fitted on these subsets of CPs. Another approach is to view the images as pieces of rubber sheet on which external forces stretching the images and internal forces defined by stiffness or smoothness constraint are applied to bring them into alignment with the minimal amount of bending and stretching [1].

Different transformations are shown in figure 1.

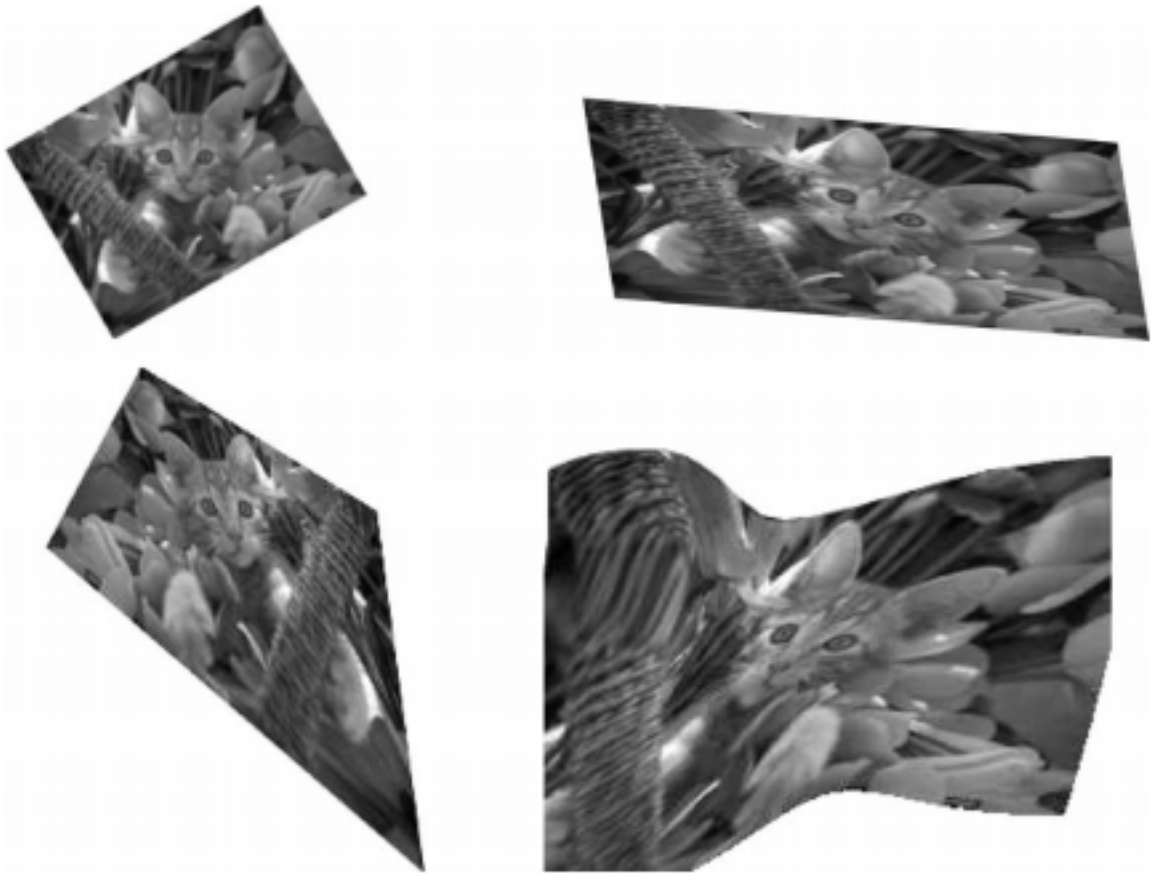

Figure 1: Examples of various mapping functions: similarity transform (top left), affine transform (top right), perspective projection (bottom left), and elastic transform (bottom right). Taken from [1]

1.3 ORB

ORB is a registration method based on features detection. The algorithm is composed of two previous works: **FAST** and **BRIEF**. FAST is a method for finding keypoints in a picture and more specifically corners. BRIEF is a feature descriptor trained on a set of 500 keypoints to find correlations between them. The authors improved them adding an orientation analysis. They showed that ORB is faster than SIFT while performing as well [3].

1.3.1 FAST and BRIEF

FAST is one of the finest methods to use when it comes to real-time keypoints search. In order to find corners, FAST takes a circle centred at a pixel and compare its intensity with the pixels within the circle. The radius of this circle is to be chosen. The authors choose a radius of 9 pixels. The authors found that FAST has too large responses along edges.

To overcome this issue, the authors employ a Harris corner measure [4] in order to select the top N keypoints. Combining Harris corner detection and FAST, a threshold is fixed in order to have more than N keypoints with FAST, then using Harris corner, the method only keep the N best points.

The BRIEF descriptor uses binary tests between pixels of a keypoint. It compares pixel intensity $\mathbf{p}$ and return 0 or 1.

$$\tau(\mathbf{p};\mathbf{x},\mathbf{y}) := \begin{cases} 1 & : \mathbf{p}(\mathbf{x}) < \mathbf{p}(\mathbf{y}) \\ 0 & : \mathbf{p}(\mathbf{x}) \geq \mathbf{p}(\mathbf{y}) \end{cases} \tag{9}$$

The feature of a keypoint is now defined as :

$$f_n(\mathbf{p}) := \sum_{1 \leq i \leq n} 2^{i-1} \tau\left(\mathbf{p};\mathbf{x}_i,\mathbf{y}_i\right). \tag{10}$$

Similarity tests use these feature with statistical tests. The author tried a few and chose the Gaussian Distribution around the centre of the pixel patch.

1.3.2 Oriented FAST and Rotated BRIEF

Keypoints detectors usually include an orientation operator, however FAST does not. To overcome this problem, the authors define a measure of corner orientation around keypoints. To do so, the moment of a patch (understand a window) of pixels is defined by Rosin [5] :

$$m_{pq} = \sum_{x,y} x^p y^q I(x,y) \tag{11}$$

and with these moments we may find the centroid:

$$C = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}}\right) \tag{12}$$

We can construct a vector from the corner's center, O, to the centroid, $\overrightarrow{OC}$. The orientation of the patch then simply is:

$$\theta = \operatorname{atan}2\left(m_{01}, m_{10}\right) \tag{13}$$

atan is the quadrant-aware version of arctan. To make this measure consistent with FAST, the authors make sure this moment is computed only in the circle used to find the corner. θ is the orientation of the corner.

BRIEF is sensible to in-plane rotation, the authors introduce Steered BRIEF. It uses the orientation of a keypoint previously computed, and the feature vector defined for the classic BRIEF. Steered BRIEF becomes :

$$g_n(\mathbf{p},\theta) := f_n(\mathbf{p}) \mid (\mathbf{x}_i,\mathbf{y}_i)$$

However, the feature vectors lose variance, due to binary tests, thus in information. Low variance makes features less discriminative. Another issue is the correlation between the test. Since corner may have a similar feature vector or be in close areas, the tests might be correlated.

To remedy this situation, the authors trained a learning method for choosing the best subset of binary tests. They used a set of 300k keypoints on the PASCAL 2006 image set. Combining, steering BRIEF and with the best subset of binary tests the authors created rBRIEF.

2 Application on leaves images

In this section, we will use the previously explained methods on the image data set given by *INRAE*. These images were preprocessed in order to keep only the foreground, which is the leaves. There are 9 rapeseed plant varieties. For each plant, 3 photos are taken. The first photo is taken 7 days after the inoculation of a fungus, then 11 days later and finally 13 days later. Our goal is to take the first picture as the reference and make the registration of the 2 others.

We will explain how we evaluate the accuracy of the registration and then show the performances of the different methods.

2.1 Metrics of evaluation

In this paragraph, we will explain why we use metric, what metric we use, and why we use specifically these metrics. Metrics are commonly used to measure the quality of a statistical model or machine learning model.

There are several ways of evaluating image registration depending on the type of error. There are localization errors which are due to slight mislocalization of the CPs. Matching errors are due to a false match. These two errors can be avoided by optimizing feature detection and consistency check.

In our case, we used two metrics to evaluate the quality of registration methods. These are the Root-Mean-Square Error (RMSE) and Structural Similarity (SSIM).

2.1.1 Root-Mean-Square Error

The root mean square error (RMSE) is the simplest complete metric and the most widespread. It is calculated by averaging the square of the intensity differences between the pixels of the reference image and the processed image. Let X be the reference image of size $M.N$

and $\hat{X}$ the resulting image (so of the same size). The RMSE formula is given by :

$$RMSE(X, \hat{X}) = \frac{1}{MN} \sum_{i}^{N} \sum_{J}^{M} |X_{i,j} - \hat{X}_{i,j}|^2 \qquad (14)$$

With $X_{i,j}$, $\hat{X}_{i,j}$ the respective pixels of the image X and $\hat{X}$.

A low RMSE value means a good quality of the registration method.

2.1.2 Structural Similarity

The $SSIM$ metric is used to measure the similarity of structure between two images, in other words, the quality of one image compared to another.

The $SSIM$ metric between two images X and Y is given by:

$$SSIM(X, Y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_x\sigma_y + c_2)(cov_{xy} + c_3)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)(\sigma_x\sigma_y + c_3)} \qquad (15)$$

with :

- μ_x the average of the pixels of the X image;
- μ_y the average of the pixels of the image Y;
- σ_x^2 the variance of the pixels of the image X;
- σ_y^2 the variance of the pixels of the image Y ;
- cov_{xy} the covariance of the pixels of the image X and Y ;
- $c_1 = (k_1 L)^2$, $c_2 = (k_2 L)^2$ and $c_3 = \frac{c_2}{2}$ are three variables whose role is to stabilize the division when the denominator is very low ;
- L the dynamics of the pixel values, i.e. 255.

The closer the SSIM value is to 1, the better the quality is.

2.2 Performances

In this section, we present the performance of the 3 registration methods.

The plant image database provided by INRAE contained a total of $n = 178$ plant images observed over three days. For reasons of memory and computing time which are increasing geometrically according to the number of images, we chose to work with a sample of plants of size $n = 35$. From a statistical point of view, the Central Limit Theorem guarantees the significance of our results and a small difference compared to the results of the total image base.

After running the algorithms of the 3 methods on the images of 35 plants and calculating the different comparison indicators, we obtained the following results shown in table 1.

We notice that each of the three methods is advantageous according to a criterion. On the basis of the mean execution time criterion, the ORB method is the most efficient, but on the basis of the mean SSIM criterion the Mutual information method is the most efficient and finally, the Coherent Point Drift method is the most efficient according to the mean square error criterion.

Overall, the most efficient method when looking at the highest SSIM/RMSE ratio is the Coherent Point Drift method.

	Average RMSE	**Average SSIM**	**Average execution time**
ORB	117.8819	0.7688	2.8185
Mutual Information	51.0177	0.9981	1736.3432
Coherent Point Drift	26.4103	0.9765	487.8795

Table 1: Performances and comparison of registration methods.

Conclusion

We have studied and compared three registration methods, namely: the ORB method, the Mutual information method, and the Coherent Point Drift method.

The implementation then a comparison of these three methods revealed that the best performing method according to the RMSE is the **Coherent Point Drift method**, whereas according to the SSIM the best method is the **Mutual Information method**. The one with the lowest average execution time is the **ORB method**.

Overall, we would choose the Coherent Point Drift method to make the registration of our images, because CPD has the lowest RMSE and a high SSIM close to the almost perfect SSIM of Mutual Information. However, we could criticize the execution time.

As a perspective, we propose to extend the study on several methods, possibly including other performance metrics such as Picture Quality Scale (PQS), Peak Signal-to-Noise Ratio(PSNR), Subjective Statistical Tests, Distortion (Mean Absolute Error), and Human Visual System (HVS).

Lexicon

- **Affine transformation** : A special set of transformations in euclidean geometry that preserve some properties of the construct being transformed.
- **Bayes' theorem** : It describes the probability of an event, based on prior knowledge of conditions that might be related to the event. Mathematically stated : $P(A \mid B) = \frac{P(B|A)P(A)}{P(B)}$, where A and B are events.
- **Central Limit Theorem** : The central limit theorem states that if you have a population with given mean and standard deviation and take sufficiently large random samples from the population with replacement, then the distribution of the sample means will be approximately normally distributed.
- **Computational cost** : is the amount of resources it requires to run an algorithm on a computer.
- **Computer vision** : is an interdisciplinary field of science that deals with how computers can acquire high-level understanding from digital images or videos.
- **Control point** : is an element of a set of points used to determine the shape of a curve, a surface or an object.
- **Corner detection** : is an approach used in computer vision to extract certain types of features and infer the content of an image.
- **EM algorithm** : It is an iterative strategy used when it is difficult to compute the likelihood because of missing data. The key idea is to compute the conditional distribution of what is missing given what is observed.
- **Entropy** : is the amount of information contained in or delivered by an information source.
- **Gaussian mixture model** : is a probabilistic model that superposes a finite number of Gaussian distributions.
- **Image alignment** : is the process of finding spatial mapping, i.e. the elements of one image which corresponds the best with the elements of a second image.
- **Image descriptors**: are descriptions of the visual features of the contents in images that produce such descriptions.
- **Image feature** : is a piece of information about the content of an image.
- **Image keypoints** : are spatial locations or points in the image that define what is interesting or what stands out in the image.
- **Image matching** : The act of checking a distance function between two sets of points which belong to some images taken from the same content.
- **Inoculation** : Action of giving a weak form of a disease to a person, an animal or a plant, usually by injection. In our case, it has been done on plants.
- **Likelihood** : The likelihood in statistics is defined as the probability of observing the data that has been observed.
- **Luminance** : is the signal that determines the contrast values of an image.
- **Mapping functions** : are functions used to apply two-dimensional textures to surfaces.
- **Pixel** : is the basic unit for measuring the luminance of a digital image.
- ***Posteriori* probability distribution** : In statistics, the posterior probability of a random variable is the conditional probability of this variable obtained after taking into account some relevant evidence.

- **Registration** : is defined as the process of establishing correspondences between two images.
- **Regularization** : is the process of adding information in order to solve an ill-posed problem or to prevent overfitting.
- **Robustness** : By "Robust Algorithms", we mean those algorithms which have the ability to deal with noisy data and also new images.
- **Signal** : is a quantity whose variation over time carries information from a source to a destination.

References

[1] Barbara Zitova (2003), *Image registration methods: a survey*, available on *https://doi.org/10.1016/S0262-8856(03)00137-9* (last visited on 07/01/2021).

[2] Andriy Myronenko and Xubo Song (2009), *Point Set Registration: Coherent Point Drift*, available on *https://arxiv.org/pdf/0905.2635.pdf* (last visited on 07/01/2021).

[3] Ethan Rublee, Vincent Rabaud, Kurt Konolige and Gary Bradski (2011), *ORB: an efficient alternative to SIFT or SURF*, available on *https://www.researchgate.net/publication/221111151_ORB_an_efficient_alternative_to_SIFT_or_SUR* (last visited on 07/01/2021).

[4] C. Harris and M. Stephens (1988), *A combined corner and edge detector*, available in *Alvey Vision Conference*, pages 147–151.

[5] P. L. Rosin (1999), *Measuring corner properties*, available in *Computer Vision and Image Understanding* 73(2):291 - 307.