Application of Low Discrepancy Sequences and Classical Control Strategies for Image Registration

Dinesh Nair
National Instruments
11500 N. Mopac B, Austin, TX 78759 USA
Email: dinesh.nair@natinst.com

Lothar Wenzel
National Instruments
11500 N. Mopac B, Austin, TX 78759 USA
Email: lothar.wenzel@natinst.com

Abstract

Image registration is the process of estimating the affine transformation made up of a rotation, a scale change, and a translation that maps one $n$-dimensional image into a second $n$-dimensional image. In practice, $n$ is usually 2 or 3. Image registration usually plays the part of an important and critical first step in applications involving the fusion of information from multiple modalities.

In this paper we introduce two important concepts: (1) non-standard quasi-random sampling of $n$-dimensional images using low discrepancy sequences to select a set of $k$ $n$-dimensional points in the image, and (2) the adaptation of a PID control strategy that uses the extracted subset of points to accurately determine the affine transformation that registers one image with another.

I. Introduction

Image registration is the process by which the correspondence between all points in two or more images of the same scene is determined. Image registration is used in image analysis tasks such as motion or change detection, fusion of data from multiple sensing modalities, and image geometric correction. There has been a tremendous increase in the need for good image registration techniques due to the increased use of temporal and multimodal 2-D and 3-D images in medical, remote sensing, and industrial applications. The aim of this paper is to introduce a couple of novel ideas for solving image registration where the images to be registered are an affine transformation, consisting of a spatial shift, rotation and scale change, apart. The registration method introduced here relies on classical control based strategies to determine the affine transformation parameters and the use of a unique sampling technique to perform the registration operation only on a subset of image points, thereby reducing the computation time.

The rest of the paper is organized as follows. We begin by introducing the theory of classical control and explain how this theory can be extended for image processing. Next, we introduce low discrepancy sequences and illustrate its use in image analysis applications. We end this paper by describing the registration framework that uses classical control theory and low discrepancy sequences in a complimentary manner. The objective of this paper is to introduce the ideas and illustrate the potential of the ideas through simple examples.

II. Classical Control

Control is an extremely well developed theory with a wide range of practical applications. Usually, this describes only the most straightforward control problems; a system is given with a specific input – output structure. The system could be a plant, an engine, a biological object, or any other structure...
with a clear input – output behavior. The number of input – output channels is low or moderate. The inputs can be interpreted as control actions on the one hand, and noise sources on the other hand. Usually, one has information about statistical properties of the disturbances. The outputs of the given system consist mainly of measurements. These measurements describe internal states, which, in most cases, are not directly measurable. In the simplest case, a so-called set point must be reached and stabilized based on appropriate control actions. Even in highly nonlinear situations, linear models can successfully be used to determine such control strategies.

The general continuous linear model

\[
\frac{dx(t)}{dt} = Ax(t) + Bu(t) \\
y(t) = Cx(t) + Du(t)
\]  

(1)

can be described by:
- matrices A, B, C, and D
- vector \( x(t) \) of internal states
- vector \( u(t) \) of control actions
- vector \( y(t) \) of measurements.

Additionally, noise terms can be added if necessary. The most successful and very common control strategy is simply a feedback according to

\[ u(t) = -K(t)x(t) \]  

(2)

The design of a valid control system consists mainly of the construction of matrices \( K(t) \) making the controlled system as efficient as possible. In typical applications the matrix \( K(t) \) is constant which simplifies the algorithms considerably.

PID controllers form a widely used subclass of control systems. More complicated systems are controlled by cascaded PID controllers. The term PID stands for a combination of \( P = \) proportional, \( I = \) integration, and \( D = \) differential components. Let \( e(t) \) be the difference between the set point and the current measurement. Clearly, the goal is to reach \( e(t) = 0 \). The PID controller reacts with a control value \( u(t) \) according to (many other notations are used)

\[ u(t) = Pe(t) + I \int_0^t de(s) + D \frac{de(t)}{dt} \]  

(3)

The quality of a PID controller depends completely on the quality of the choice of parameters \( (P, I, D) \). The preferable strategy is to tune (auto-tune) the PID controller using and observing the real behavior of the given system.

2.1 Control and Image Analysis

At a first glance, images (at least beyond a certain size) don’t match the requirements of classical control. Each pixel can be interpreted as a separate channel. Though these channels are not completely independent, the overall degrees of freedom are enormous. On the other hand, tasks such as object tracking, the determination of sub-pixel accuracy in the final phase of pattern matching algorithms, or image registration can be formulated in terms of classical control theory. Here, we will closely follow Hoger and Belhumeur’s approach. In [1] they describe iterative algorithms capable of tracking objects under very general motion models of the target (see also [2]). A straightforward reformulation and redefinition of the goal leads to a registration algorithm. Three new features are added.

First, Hoger and Bulhumeur’s algorithms are tracking mechanisms, whereas our strategy tries to register two given and similar images. Second, the method developed in [1] is a direct consequence of a certain minimization routine. The result can be interpreted as a control strategy with \( P=1, I=0, \) and \( D=0 \). Third, to reduce the computational load, our algorithm is based on some well-defined parts of given images. More precisely, instead of taking into account all pixels in the image, we restrict the computation to low discrepancy sets. In doing so, there is no significant loss of accuracy but a speed up that is in the order of
pixels used from the images to be registered.

III. Low Discrepancy Sequences

Pseudo-random sequences have been used as a deterministic alternative to random sequences for use in Monte Carlo methods for solving different problems. Recently, it was discovered that there is a relationship between low discrepancy sets and the efficient evaluation of higher-dimensional integrals. Theory suggests that for midsize dimensional problems, algorithms based on low discrepancy sets should outperform all other existing methods by an order of magnitude in terms of the number of samples required to characterize the problem.

Given a function \( f(x) \) the problem of calculating the integral
\[
I(f) = \int_0^1 f(x) dx
\] (4)
in the most efficient manner is not a well-posed problem. An approximate strategy could be based on the following procedure:

(A) Construct an infinite sequence \( \{x_1, x_2, x_3, ..., x_n, \ldots\} \) of real numbers in \([0, 1]\) that does not depend on a specific function \( f \) (we do not know anything about \( f \) in advance, except some general smoothness properties).

(B) During the \( n^{th} \) step of the algorithm calculate \( f(x_n) \) and the approximation to the integral in (4) as:
\[
I_n(f) = \frac{f(x_1) + ... + f(x_n)}{n}
\] (5)
If a certain criterion is satisfied stop, else repeat step (B). The stopping criterion depends strongly on objectives such as accuracy or speed.

How does this algorithm differ from standard methods such as trapezoidal rule which is based on equally distributed points in \([0, 1]\)? First, there is no relationship between consecutive sets \( x_i(n) = i/n \) and \( x_i(n) = i/(n+1) \). In other words, if the approximation given in equation (5) fails a given goal, a complete recalculation of numerous \( f \)-values is necessary. On the other hand, it is well known that the trapezoidal rule gives a \( 1/n^2 \) rate of convergence for a given continuous function \( f \).

Obviously, the quality of the trapezoidal rule is based on a highly homogeneous set of points. To quantify the homogeneity of a finite set of points, the definition of the so-called discrepancy of a given set was introduced ([3], [4]):
\[
D(X) = \sup_R |m(R) - p(R)|
\] (6)
Here, \( R \) runs over all rectangles \([0, r]\) with \( 0 \leq r \leq 1 \), \( m(R) \) stands for the length \( r \) of the closed interval \( R \), and \( p(R) \) is the ratio of the number of points of \( X \) in \( R \) and the number of all points of \( X \). The definition given in equation (6) can be generalized to the case of \( d \) dimensions (\( d=2, 3, \ldots \)), where the term interval must be interpreted as an \( d \) dimensional rectangle. The lower the discrepancy the better or more homogeneous the distribution of the set. The discrepancy of an infinite sequence \( X = \{x_1, x_2, ..., x_n, \ldots\} \), is a new sequence of positive real numbers \( D(X_n) \), where \( X_n \) stands for the first \( n \) elements of \( X \). Other definitions for the discrepancy do exist that avoid the worst-case scenario according to (6).

Clearly, there exist a set of points of given length, that realizes the lowest discrepancy. It is well-known ([5]) that the following inequality holds true for all finite sequences \( X \) of length \( n \) in the \( d \) dimensional unit cube
\[
D(X) \leq B_d \left( \frac{\log n}{n} \right)^{(d-1)/2}.
\] (7)
\( B_d \) depends only on \( d \). Except for the trivial case \( d=1 \), it is not known whether or not the theoretical lower bound is attainable.

Many schemes to build finite sequences \( X \) of length \( n \) do exist that deliver a slightly worse limit
\[ D(X) \leq B_d \left( \frac{\log n}{n} \right)^d. \]

There are also infinite sequences \( X \) with
\[ D(X_n) \leq B_d \left( \frac{\log n}{n} \right)^d \]
for all natural numbers \( n \). The latter gave rise to the definition of so-called low discrepancy (infinite) sequences \( X \). The inequality in equation (9) must be valid for all \( n \), where \( B_d \) is an appropriately chosen constant. Low discrepancy sequences are also known as quasi-random sequences.

On average, a randomly chosen sequence in \([0, 1]\) has a discrepancy value in the order of \( 1/\sqrt{n} \) which is far beyond the low discrepancy value in the order of \( (\log n)^d/n \).

The relationship between the integration in equation (4), its \( d \)-dimensional generalization, and the approximation given by equation (6) for an infinite sequence \( X = \{x_1, x_2, ... \} \) is given by the Koksma-Hlawka inequality.
\[ \| I(f) - \int f(x) \| \leq V(f) \cdot D \]
where, \( V(f) \) is the variation of the function in the sense of Hardy and Krause. Even if a finite \( V(f) \) exists (e.g., for smooth functions), the inequality in equation (10) is of no practical use. Very often, the real value of \( V(f) \) is unknown and describes only the worst-case. But, at least in principle, a low discrepancy set \( X \) should be preferred to all other sequences.

3.1 The Halton Low Discrepancy Sequence

Many of the well-studied low discrepancy sequences in the \( d \)-dimensional square can be constructed as combinations of 1-dimensional low-discrepancy sequences. The most popular low discrepancy sequences are based on schemes introduced by Richtmeyer [5], Halton [4], Sobol’ [6, 7], Niederreiter [8], and Faure [9]. The book [7] gives a comprehensive introduction into the implementation of low discrepancy sequences (Halton and Sobol’).

We will explain the Halton method here in detail. All of our test results are based on Halton and Sobol’ sequences.

Halton sequences in 1-d start with the choice of a natural number greater than 1. Though not absolutely necessary, prime numbers \( p = 2, 3, 5, ... \) are typically chosen. If \( p \) is a given prime number and \( x_n \) the \( n \)th element of the Halton sequence, the following algorithm determines \( x_n \).

(A) write \( n \) down in the \( p \)-ary system
\[ n = n_q \ldots n_0, n = n_0 + n_1 \cdot p + \ldots + n_q \cdot p^q \]
(B) Reverse the order of the digits and add the \( p \)-ary point
\[ 0.n_p n_q \ldots n_0 \]
(C) It is
\[ x_n = n_0 \cdot p^{-1} + n_1 \cdot p^{-2} + \ldots + n_q \cdot p^{-(q+1)} \]

The \( n \)th element of the Halton sequence can be calculated independently of all other elements. As mentioned above, in \( d \) dimensions one has to interpret different 1-dimensional Halton sequences as coordinates of points in \( d \) dimensions. It is very common to start with the first \( d \) prime numbers.

Figure 1 shows the first 100 elements of a Halton sequence in the unit square for two different valid choices of starting prime numbers \((2, 3)\) and \((13, 17)\). Obviously, the first couple performs much better at least at the very beginning of the sequence. Because of the relatively low number of pixels in typical image processing applications, homogeneity (low discrepancy value) is always desirable. That is why, all experiments were based on the most straightforward combination \((2, 3)\).
Halton sequences in 1-D are low-discrepancy sets in the sense of equation (8). More precisely, for all \( n \) and for all Halton sequences \( X \) that are based on a prime number \( p \) it is (5)

\[
D(X) \leq B \frac{\log n}{n} \quad \text{with}
\]

\[
B = \begin{cases} 
\frac{p^2}{4(p+1) \log p} & \text{when } p \text{ is even} \\
\frac{p-1}{4 \log p} & \text{when } p \text{ is odd}
\end{cases}
\]

A similar result (see [5] again) holds true for Halton sequences in \( d \) dimensional unit squares. In a 2-dimensional unit square for the \((p,q)\) Halton sequence with prime numbers \( p \) and \( q \) the discrepancy is

\[
D(X) \leq \frac{2}{n} \left( \frac{\log n}{n} \right)^2 \left[ \left( \frac{p-1}{2 \log p} + \frac{p+1}{2 \log n} \right) \left( \frac{q-1}{2 \log q} + \frac{q+1}{2 \log n} \right) \right]
\]

3.2 Low Discrepancy sequences and images

Most routines in image processing are based on sampled objects where the resolution is limited by hardware and can not be influenced by the user. On the other hand, low discrepancy sequences are inherently continuous. Given a digital image of size \( N^{(1)} \times N^{(2)} \), the \( n^{th} \) pixel element according to the given low discrepancy sequence \((x_{1}, x_{2}, ..., x_{n}, ...)\) in the unit square has the coordinates \([N^{(1)}, x_{n}^{(1)}]\) and \([N^{(2)}, x_{n}^{(2)}]\), where \( x_{n} = (x_{n}^{(1)}, x_{n}^{(2)}) \) and \([\cdot]\) stands for the nearest integer. Because of the homogeneity of low discrepancy sequences, double hits are impossible if \( n \) is sufficiently small. The final goal of a combination of low discrepancy sequences and image processing is the significant reduction of information processing necessary for image analysis. Good approximations must be delivered with the aid of a small percentage of all available pixel values.

The field of image processing and image understanding can potentially take advantage of specific properties of low discrepancy sets. To illustrate this, we applied the theory of low discrepancy sequences to some relatively simple image processing and computer vision related operations such as the estimation of gray level image statistics and fast location of objects in a binary image. The results of our experiments are tabulated below. In the first experiment, we estimated the average number of points as a percentage of the mage size needed to estimate the mean of the image with a certain accuracy. Accuracy was defined in terms of the absolute difference from the true value. The image database used in this experiment comprised of images of man-made objects, textures, medical imagery, fractals etc.

<table>
<thead>
<tr>
<th>Method</th>
<th>1.0 Accuracy</th>
<th>0.5 Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Halton</td>
<td>0.2</td>
<td>0.7</td>
</tr>
<tr>
<td>Random</td>
<td>0.7</td>
<td>3.3</td>
</tr>
<tr>
<td>Grid</td>
<td>0.6</td>
<td>2.3</td>
</tr>
</tbody>
</table>

Table 1: % of number of image points required for the estimation of the mean gray-scale value of an image using different sampling techniques.

In the second experiment, the objective was to determine the number of points needed to locate a randomly placed binary rectangle in an image with probability beyond 0.5. The image size was 512x512. The object was assumed to be located if a point on the sequence fell inside the boundaries of the object.
<table>
<thead>
<tr>
<th>Method</th>
<th>Number of Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Halton</td>
<td>100</td>
</tr>
<tr>
<td>Random</td>
<td>115</td>
</tr>
<tr>
<td>Grid</td>
<td>114</td>
</tr>
</tbody>
</table>

Table 2: Number of sample points required for locating an arbitrarily placed object in a 512x512 binary image using different sampling techniques.

Our experiments show that compared to standard methods, the proposed new algorithms require fewer points than regular grid-based sampling and random sampling to accurately characterize images. Hence these algorithms are faster and statistically more robust than conventional sampling techniques.

IV. Image Registration Framework

This section presents the method used for image registration using low discrepancy sequences and a control strategy.

Given two images imageA and imageB, let imageB be a shifted, rotated, and scaled version of imageA. We assume that the four parameters \( x, y, \theta, \) and \( s \) (\( x\)-shift, \( y\)-shift, \( \theta\)-rotation, \( s\)-scaling factor) are relatively close to 0, 0, 0, and 1, respectively. Given two reasonably sized regions, regionA and regionB, where regionA is part of imageA and regionB is part of imageB respectively, regionA matches regionB with unknown values \( x, y, \theta, \) and \( s \). The goal is to determine these values accurately.

4.1 The control strategy

The goal is to reduce the distance between regionA and a shifted, rotated, and scaled version of regionB (as part of imageB) with the aid of a step-by-step approach. The original situation is described by an unknown \( x, y, \theta, \) and \( s \). The control strategy can be divided into two parts ([1]).

Preparation:
1. Calculate the Prewitt derivatives of regionA and flatten these derivatives. This results in vectors \( I_x \) and \( I_y \)
2. Build the matrix

   \[
   M_0 = \begin{bmatrix} I_x & I_y & -yI_x + xI_y & xI_x + yI_y \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}
   \]

   Calculate the matrix \( \lambda = (M_0^T M_0)^{-1} M_0^T \).

   The matrix \( M_0^T M_0 \) is 4x4 and non-singular (with exception of some pathological situations).

Control:
1. Let \((x, y, \theta, s)\) be the current estimates.
2. Let \( \rho(\theta,s) \) be the matrix

   \[
   \rho(\theta,s) = \begin{bmatrix} \cos(\theta)/s & -\sin(\theta)/s & 0 & 0 \\ \sin(\theta)/s & \cos(\theta)/s & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1/s \end{bmatrix}
   \]

3. Let \( e \) be the flattened difference (pixelwise) between regionA and the shifted, rotated, and scaled version of regionB. Then

   \[
   [\Delta x, \Delta y, \Delta \theta, \Delta s] = \rho(\theta,s)^* \lambda^* e
   \]

   and

   \[
   x_{\text{new}} = x + \Delta x \\
   y_{\text{new}} = y + \Delta y \\
   \theta_{\text{new}} = \theta + \Delta \theta \\
   s_{\text{new}} = s + \Delta s
   \]

4. Calculate the new \((x_{\text{new}}, y_{\text{new}}, \theta_{\text{new}}, s_{\text{new}})\) version of regionB, i.e. shift the old regionB by \((\Delta x, \Delta y)\), rotate it by \( \Delta \theta \), and shrink or stretch it by the factor \( \Delta s \), using bilinear interpolation.

5. Depending on the value of the norm of \( e \) stop or go to (3) again.

Figure 2 shows a typical pair of images that were
used for the registration experiments. In this example the image shown in Figure 2(b) is a shifted, rotated, and scaled version of the image in Figure 2(a). Figure 3 depicts the typical response of this registration algorithm. The \((x, y, \theta, s)\) data slowly approach the real values which are completely known in this example.

![Figure 2: A pair of images used for the registration example. The image in (b) is a shifted, rotated, and scaled version of the image in (a).](image)

![Figure 3: Response time of the control-based registration technique using the entire image.](image)

More efficient strategies:
One of the drawbacks of the described algorithm is its sluggish convergence behavior. Though (3) is a direct consequence of an optimization procedure ([1]), it is usually not the fastest procedure. The performance increases by introducing constant factors \(k_x, k_y, k_\theta, \text{and } k_s\). The new estimates are:

\[
\begin{align*}
x_{\text{new}} &= x + k_x \Delta x \\
y_{\text{new}} &= y + k_y \Delta y \\
\theta_{\text{new}} &= \theta + k_\theta \Delta \theta \\

s_{\text{new}} &= s + k_s \Delta s
\end{align*}
\]

This is nothing else but four \(P\)-controllers acting on the four channels \(x, y, \theta, \text{and } s\). Figure 4 shows the same situation as before but with \(k\)-values: \(k_x=5, k_y=5, k_\theta=5, \text{and } k_s=5\). The performance is significantly better than with the choice of \((1, 1, 1, 1)\).

![Figure 4: Response of the registration of the images in Figure 2 using a \(P\) value of 5 for each of the affine parameters to be estimated.](image)

**Autotuning:**
The last remark gives rise to the introduction of complete PID-controllers for the four above-mentioned channels. The problem is that there is no set of \(P\)-, \(I\)-, and \(D\)-components covering all situations equally well. In other words, depending on the image content different sets of parameters must be chosen. In the majority of our experiments the set

k_x = 5, k_y = 5, k\theta = 5, \text{ and } k_s = 5(\text{where } s \text{ is measured in } \% \times 100) \text{ showed an excellent behavior. The I-}

\text{and D-components were set to 0.}

If any of the images being registered is noisy, the D-values should be 0 or very close to this value. In case of noise-free images or filtered versions of the originals a D value different from 0 can result in an even more improved speed of convergence. By introducing an I term, you can reduce the oscillatory behavior seen in figure 4. In general, there is no strict method of determining the optimal sets of P-, I-, and D-parameters. A successful approach, if applicable, can be described as follows. Shift, rotate, and scale region A artificially and measure the speed of convergence under different conditions, i.e. different sets of parameters. Choose the set with the best behavior.

This method has some obvious similarities to auto-tuning that is used commonly in classical control. Another method is based on adaptive control, i.e. modifications of an initial set of parameters in dependence of the norm of the error e. We prefer the first procedure to the second one because the auto-tuning method produces very similar results for images belonging to the same family.

4.2 Use of Low discrepancy Sequence Points

A careful study of the developed control strategy reveals that the Prewitt derivatives I_x and I_y can easily restricted to specific parts of the image without changing the algorithms. Earlier we demonstrated that deterministic random sequences (low discrepancy sets) out perform other choices based on the same amount of pixels. More precisely, an excellent estimate of the average gray level of a given image can be achieved if only a very small percentage of all pixel values are considered. The well distributed Halton or Sobol’ sequences interpreted as pixel positions on an image deliver much better results than randomly chosen pixels or grid like structures. Another advantage of low discrepancy sequences is the ability to add further points without loosing results achieved so far. Both properties make Halton, Sobol’ and other low discrepancy sequences superior to comparable choices.

Figure 5 is the result of the new control strategy where only 10% of all points were used. These pixels belong to a Halton sequences. Compared to a full set of pixels, the speed of convergence is as fast as in the original case. Randomly chosen pixels can not guarantee this behavior.

5.3.1 Convergence behavior of the registration algorithm using Halton points.

The described algorithms have their limitations. They do not perform well if the shift is beyond +/-4 pixels (both in x- and y direction), if the rotation is beyond +/-4 degrees, and if the scaling factor is beyond +/-4%. Typical region sizes are in the order of 80x80. If one can not guarantee these conditions an additional step is necessary. Based on pattern matching or
similar techniques, a first estimate must be generated that satisfies the above mentioned parameters.

**IV. Summary**

In this paper a new image registration technique is presented. The problem of finding the affine transformation parameters between the images to be registered is posed as a classical control problem. An efficient and robust image registration is performed by sampling the images to be registered using low discrepancy sequences and estimating the transformation between two subsets using a strategy based on the PID control.

**References**


[9] H. Faure; Discrepance de suites associees a un systeme de numeration (en dimension s), Acta Arithmetica 41, 337-351, 1982