# Non-Rigid Image Registration

**Rhoda Baggs**
Department of Computer Information Systems
Florida Institute of Technology.
150 West University Blvd.
Melbourne, FL 32901-6975
rbaggs@fit.edu

**Dan E. Tamir**
Department of Computer Science
Texas State University
601 University Drive
San Marcos, TX, 78666
dt19@txstate.edu

**Abstract**

An important hypothesis of the research reported in this paper is that dynamic space warping (DSW), a dynamic programming (DP) technique (Bellman 1965, Brown 1996), can be effectively used for object matching and for inference of the image registration warping function. For this end, a novel, robust, and efficient method for non-rigid image registration using DSW in the "length–code" domain is developed. Length codes are affine transformation invariant object shape features (Baggs 1996). DSW is utilized to estimate linear, nonlinear, and non-rigid warping (Keogh 2006). The DP algorithm maps length codes in an attempt to identify objects that appear in more than one image and uses their coordinates to register the images.

## Introduction

This paper shows that under reasonable constraints and assumptions, dynamic space warping (DSW), a dynamic programming (DP) technique (Bellman 1965), can be applied successfully to the object matching and image registration problem (Alajani 2006). The DSW is applied to a set of affine-transform-invariant object-shape-features referred to as length code (Foley 1990, Keogh 2006, Shabi 2001).

Misalignment between images of the same scene is often the result of differences in image acquisition conditions or acquisition equipment. A typical effect of the acquisition differences is non linear geometric distortion (or warping). For example, aerial photography of the same area taken by two different airplanes may introduce different (non-linear) scaling, translation, rotation, and shearing to the obtained image objects. In addition, these differences may affect the light intensity, and image background.

As stated by Haralick and Shapiro (Haralick 2002), "Image registration is the geometric warping process by which two images of different geometries but of the same set of objects are spatially transformed so that the size, shape, position, and orientation of any object on one image are made to be the same size, shape, position, and orientation of that object on the other image." An overview of registration techniques can be found in (Zitova 2003).



Hence, the goal of image registration is to identify and apply the transformation function that aligns the images. This function represents an approximation of an optimal mapping of pixels from one image onto the other.

In this paper it is assumed that the geometric distortion can be represented to a good degree of approximation using affine transformations. In other words, we are searching for the best affine transformation that can approximate the actual distortion between the images. Except for normalization of brightness levels, the algorithms do not take into account potential differences in light conditions and background changes.

Dynamic time warping (DTW) has been thoroughly investigated and implemented in speech recognition systems (Itakura 1975). Dynamic space warping is much less explored. Only a few papers on DSW applications are available (Shabi 2001).

Keogh extends DTW to DSW and uses it for matching objects in the signature domain (Keogh 2006). While he takes into account translation and rotation invariance, he is not concerned with invariance with respect to the entire affine transformations. The advantage of the features presented here is that they are more robust, computationally efficient, and allow for efficient introduction of scaling invariance. In addition, Keogh is not concerned with the application of DSW to image registration.

This is the first contribution in the area of image processing which applies DSW to signatures of images for the purposes of object matching within the context of the image registration problem. In particular, the signature produced for objects in an image is the length code algorithm developed by Baggs (Baggs 1997).

### The registration process

Two images are involved in the registration process the "reference image" and the "inspected image." The process described in this paper includes identifying matches between objects in the inspected image and the reference image. Unique object points are used to estimate the warping transformation function. The function is used to map pixels of the inspected image to pixels in the reference image.

The registration process reported in this paper involves the following steps:

- Identify all the objects in the reference image, and all the objects in the inspected image,
- Extract shape features (length codes) of objects.
- Implement the DSW pattern matching algorithm to match objects from the reference image with objects from the inspected image.
- Pick the K best object matches and store each match as an ordered pair of the objects centroids.
- Using the centroids, compute and apply a transformation function which registers the images.

We add more details to these procedures in respective sections of the paper. The Object Matching phase is the primary focus of this paper.

### Experiments

A set of experiments has been performed in order to validate the utility and robustness of the proposed method.

The results of the experiments, reported in the section “Experiments and Results,” show that very accurate matching with low error rates are achieved and produce high quality registration.

## The Length Code Registration Algorithm

The object matching phase utilizes shape features referred to as length codes. In order to obtain length codes the image has to be segmented, objects have to be identified, and the pixels that reside on object contours have to be marked (Alajan 2006, Brown 1992).

### Object Identification

Image segmentation can be implemented in several different ways (Haralick 2002). The method chosen is the robust vector quantization (VQ) image segmentation method described in (Coleman 1979). Moreover, our implementation of the VQ algorithm (Tamir 2007) is about four times faster than the commonly used Linde, Buzo, and Gray (Linde 1980) LBG algorithm. This results in a computationally efficient segmentation method.

One advantage of the VQ is that it is less sensitive to changes in the lighting of a scene between two consecutive image acquisitions. These changes may affect the brightness level of an object that appears in both images, distort its shape, and affect the probability of matching the two instances of the object. To further reduce the sensitivity of the segmentation to light changes we normalize the gray levels of the reference and the inspected images before applying the VQ-based segmentation algorithm. Following the segmentation, a connected component labeling (CCL) and a contour following algorithm are applied (Haralick 2002).

### Shape Representation Using Length Code

Object shape features such as chain codes, Fourier descriptors, and object-signature are often used for object matching and image registration (Brown 1992, Foley 1990, Haralick 2002). A desirable characteristic of many of the shape features is their set of invariants which can provide matching robustness. For instance, chain-codes are invariant to translation.

This work uses a variant of the **object signature** referred to as the “length code”. The term object-signature is “overloaded.” In the context of this paper it is the set of distances measured from a fixed-point referred to as the ‘center’ (generally this point is centroid of the object) to pixels that reside on the object boundary (contour). There are two ways to produce the set of distances. In a fixed angle length code, the distances are measured in increments of equal angles. A second approach is to measure the distance from the center to every pixel on the contour of the object. A problem with the fixed angle method occurs with concave objects. The same angle may yield more than one distance to points on the contour. The “every-pixel” method removes the requirement for convexity and can use an arbitrary center. Due to the invariance toward convexity the “every-pixel” method is used in this paper.

Length coding has several advantages over other shape representation methods with respect to object matching. The main advantage is that it is less sensitive to “noise”. This is due to the fact that a small change in shape, which may be due to noise, does minimal change to the distance between the centroid and contour points. On the other hand, since chain codes are three bit numbers, then a small change in the boundary can result in a significant change (relative to the range of [0,7]) in the chain code. The same applies to Fourier descriptors which are based on the values of the tangents to the contour at the contour points. A small change in shape at a given boundary pixel can result in selecting a different member of the 8-neighbors of that pixel thereby producing a relatively large change in the sequence of tangents and their Fourier transform. In addition, length coding is invariant to the entire set of affine transformations. Finally, this representation converts a two-dimensional problem into a one-dimensional problem without loss of information and provides a computationally efficient framework.

#### Length Coding

The cyclic auto correlation function of the signature sequence is rotational and translational invariant (Baggs 1996, Haralick 2002). The first autocorrelation coefficient of the signature, $R_0$, is the sum of squares of the values of signature elements. Hence, it approximates the energy of the signal and is proportional to the area of the object. Thus, in order to normalize the sequence each autocorrelation coefficient $R_i$ is divided by the coefficient $R_0$. The resultant sequence is scale invariant. It is referred to as the "length code" of an object.

## The DP Object Matching Algorithm

An image transformation is a mapping of the pixel locations in one image to a new location in a second image. The transformation is assumed to be global. That is, the same transformation equation is applied to all the pixels in the image.

The set of affine transformations accommodate planar mappings, i.e., it preserves parallel lines and equi-spaced points (Foley 1990). Translation, rotation, and scaling are the three basic affine transformations. This paper assumes that the warping between the reference image and the inspected image can be approximated by an affine transformation applied to the reference image. Hence, the goal of the inference of the image transforms stage is to identify the inverse transformation (the transformation that maps the inspected image to the reference image), and apply the inverse transformation to the inspected image.

The coordinates of at least three points in the reference image along with the coordinates of the compatible three points in the inspected image are required in order to compute the warping transformation. If only three pairs of points are available, then a set of six equations in six unknowns is inferred and used to compute the transformation. Generally, more than three pairs of points are available producing an over determined set of linear equations. The least square method is used to solve these equations. This is further detailed in section 2.5.

A normalized homogeneous coordinate space is used (Foley 1990). In this space, the three basic transformations are represented by 3x3 matrices. Arbitrary transformations, which are the result of a sequence of basic transformations, are computed through 3x3 matrix multiplications and yield a 3x3 matrix. The actual image transformation is computed by multiplying the homogeneous coordinates of pixels by the transformation matrix. Let $P=[i,j,1]$ be the homogenous coordinates of a pixel in the reference image. Then, $Q=[m,n,1]$ are the coordinates of this pixel after applying affine transformation to the image. See (EQ-1):

$$[m,n,1]=[i,j,1]\times A=[i,j,1]\times\begin{bmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & 1 \end{bmatrix} \qquad \text{(EQ-1)}$$

Shearing is not a basic transformation. It can be obtained through a sequence of basic transformations such as rotation, and scaling. Nevertheless, we dedicate special consideration to shearing as it is a typical warping that occurs in digital image processing.

It may appear that the length code is invariant only with respect to affine transformations performed in a coordinate system where the origin is the object's center. This is not the case, since an affine transformation with respect to arbitrary point '$C_1$' can be expressed as a combination of affine transformations where the origin is the point '$C_2$'.

### Inferring the Inverse Transform

Consider the points *P*, *Q* and the affine transformation that maps *P* into *Q* (say $A$), then the inverse transformation exists, is itself affine, and defines *P* in terms of *Q*. Given a transformation matrix $A$, then its inverse $A^{-1}$ is defined by: $A^{-1}=(adj(A)/\det(A))$. The adjoint (*adj*) of a matrix is the transpose of the matrix of cofactors and '*det*' stands for the determinant of a matrix (Baggs 1997). The inverse transformation is then defined by the equation (EQ-2):

$$P=[i,j,1]=[m,n,1]\times A^{-1}=[m,n,1]\times\frac{1}{a_{11}a_{22}-a_{12}a_{21}}\times$$
$$\times\begin{bmatrix} a_{22} & -a_{12} & 0 \\ -a_{21} & a_{11} & 0 \\ a_{21}a_{32}-a_{31}a_{22} & a_{31}a_{12}-a_{11}a_{32} & a_{11}a_{22}-a_{12}a_{21} \end{bmatrix} \quad \text{(EQ-2)}$$

Thus, an affine transformation has six degrees of freedom, relating directly to coefficients $\{a_{ll}, a_{21}, a_{31}, a_{12}, a_{22}, a_{32}\}$. Hence, six equations have to be formulated and solved in order to uniquely determine the transformation. This can be accomplished by identifying the location of three non-collinear points in both images (reference and inspected). Let $P_1$, $P_2$, and $P_3$, $Q_1$, $Q_2$, and $Q_3$ be these three points in the reference and inspected images, respectively, EQ-2 expresses their relationship in the form of a matrix equation. The six unknown coefficients of the affine mapping are determined by solving the system of six linear equations contained in EQ-2 (Baggs 1997).

When more than three correspondence points are available, and when these points are known to contain errors, it is common practice to approximate the coefficients by solving an over determined system of equations. In that case, *A* is no longer a square 3x3 matrix and it must be inverted using any technique that solves a least-squares (LS) linear system problem. The reader is referred to (Baggs 1997) for a discussion on this topic.

Let *K* be the number of pairs of matching objects identified in the reference and inspected images. We consider four cases: 1) $K=0$, in this case the registration is deemed to be impossible. 2) $0<K<3$, in this case we use the obtained matches to generate three matching points. These points are the centroids and unique points that belong to the object's contour (e.g., the points with maximum distance from the centroids). 3) $K=3$, we use the three pairs of centroids as the 3 points. Finally, 4) $3<K$, we solve an over-determined equation to identify the transformation. We discuss the option of using more than

three points when $0 < K \le 3$ (cases (2) and (3) above) in section 4.0.

## Experiments and Results

Two types of experiments are executed. The first set of experiments includes synthetic data and the second includes real data, i.e., natural images obtained through image acquisition devices.

The input to each experiment is a reference image *R* with $N_r$ objects and an inspected image *I* with $N_i$ objects. In most of the experiments, it is assumed that the image *I* contains the same (or very similar) scene as the image *R*. The difference between the two images is due to the fact that they are obtained through different acquisition devices, conditions, perspectives, and times. The result of these differences is warping. It is further assumed that the warping can be approximated by a set of affine transformation applied to the image *R*. The experiments are all non-supervised and include the following steps:

- Preprocessing
    - Segmentation
    - Connected component labeling (CCL)
    - Contour following
    - Length code extraction
- DSW matching
- Inferring and applying the inverse transformation
- Obtaining a difference image and measuring the registration quality.

The pre-processing stage identifies objects in the images and represents these objects using length code. The DSW matching stage produces a dissimilarity matrix. The matrix element $[d_{r,i}]$ is the distance between object $O_R$ from the image *R* and object $O_I$ from the image *I*. The inference stage uses $3 \le K$ unique points identified in pairs of matching objects in order to estimate the warping function (the inverse transformation). This function is then applied to the pixels of the inspected image.

Experimental results are measured by the recognition accuracy and the recognition error rates. In addition, the quality of the results is evaluated using the power of the difference image along with the power of the reference image. The ratio of the powers is used to calculate the signal to noise ratio (SNR). High SNR demonstrates that the inverse transformation has accurately mapped the inspected image to the reference image.

In the case of RGB images the L2 norm used to compute the SNR is computed in the three dimensional RGB space. To further elaborate, consider the image $T$. That is, the image obtained from the image *I* after applying the inverse transformation $A^{-1}$ to the image *I*. This is the best approximation of the original image *R* obtained through the proposed algorithms. The difference image $D$ is constructed through pixel wise absolute-value subtraction of the corresponding pixels in the images $R$ and $T$. The SNR value is computed using the following formulation:

Let $\sigma_R^2 = \sum_{P \in R} P^2$ and $\sigma_D^2 = \sum_{Q \in D} Q^2$, then the SNR (measured in DB) associated with the image *D* (and hence with the experiment of matching the image *I* to the image *R*) is given by: $SNR_D = 10 \log_{10} \frac{\sigma_R^2}{\sigma_D^2}$ DB.

Note that high SNR denotes better registration quality as it means that the difference image contains a small amount of information.

### Experiments Using Synthetic Data

For the synthetic data experiments, 27 different computer generated images are used in order to supply complete control over the nature of the input. Each image consists of a background which is ignored and one set of pixels representing a “simple”, filled in, object. The types of objects created are ellipses, circles, triangles, squares, rectangles, and parallelograms. Different sizes and positioning of the objects provide suitable and careful testing of translation, rotation, uniform scaling, and non-uniform scaling, and shearing.

A total of 351 experiments were performed. We refer to an error as a ‘type-1 error’ to be the case when an object from one class (e.g., ellipsoids) did not match an object from the same class. A ‘type-2’ error occurs when an object from one class matches an object from a different class.

Table 1 shows the match matrix for the entire set of experiments with synthetic data. The table is derived from a decision function with a distance threshold of 4.0. The matching score is denoted in the percentage with respect to the number of experiments per class.

| $Th \le 4.0$ | Circle / Ellipsoid | Triangle | Rectangle |
|---|---|---|---|
| Circle / Ellipsoid | 96% | 0 | 0 |
| Triangle | 0 | 100% | 0 |
| Square / Rectangle | 0 | 8% | 100% |

Table 1, The match matrix

Table 1, presents a high recognition rate of 96% and above, and an 8% or below type-2 error rate. Type-1 error rate can be deduced from the table by subtracting 100% from each diagonal element of the table. Hence, it is 4% or below. The total error rate (type-1 and type-2 errors) over the entire set of experiments is 4.2%.

## Experiments With Natural Images

Two types of experiments were performed on natural images. First, a given image was used as the reference image. Different versions of this image, obtained through a controlled synthetic affine transformation of the image were used as the inspected images. The second includes pairs of natural images of different scenes and pairs of images of the same scene obtained from different camera positions. Each experiment yields an individual match matrix and an SNR value.

### Controlled Natural Images

All of the experiments discussed in this section involve the comparison of the image POKER (shown below) with a copy of POKER, artificially transformed.

Ten types of transformations are applied to the original POKER image including: translation, rotation, uniform and non-uniform scaling, uniform and non-uniform shearing, compositions of translation, rotation and non-uniform scaling / shearing. Within each group mentioned above, 4 different images are created, each with different values for the transformation; i.e. each represented by a different transformation matrix. Thus a total of 40 images used to compare to POKER have been constructed.

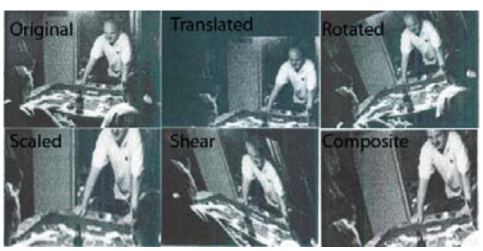


Figure 1, “Poker” artificially transformed

The applied warping is quite severe. It includes translations by up to 28 pixels, rotations by up to $20^{o}$, up to 2.25 times scaling, and up to $40^{0}$ shear. The non-uniform transformations use the same ranges with different values for the ‘*x*’ axis and for the ‘*y*’ axis. Composite transformations include combinations of uniform and non uniform rotation, translation, scaling, and shearing.

Table 2 summarizes the average SNR obtained over a class of experiments with a given set of transformations parameters (listed in the table).

| Description | Parameters | Average SNR |
|---|---|---|
| Translations | (7, 7) | 15.40 |
| Rotations | $5^{o}$ | 14.31 |
| Scaling | 1.25 | 14.21 |
| Shearing | ($10^{o}$, $10^{o}$) | 14.26 |
| Non-uniform scale | (1.25, 1.5) | 11.76 |
| Non-uniform shear | ($10^{o}$, $15^{o}$) | 13.34 |
| Compositions | See text | 11.99 |
| All Transformations | | 13.57 |

Table 2, Average SNR

A relatively large SNR indicates a better matching and a good range for SNR would be between 25 to 35 DB. The SNR obtained here is lower. This is partially due to the fact that the warping is severe and might be due to data rounding. The transformations yield real values which are rounded in order to obtain discrete pixel locations. The rounding is done twice. First, when *I* is generated from *R*. Next, when *T* is generated from *I*. As a result the accumulated error is relatively large. It should be noted, however, that visual inspection of the images shows that the registration is quite good. It is close to registration obtained via manual alignment and is sufficient in order to implement operations such as change detection.

### Arbitrary Natural Images

The experiments described in this section include pairs of natural images. A pair of images is considered as ‘like images’ if they are images of the same scene obtained from different perspective points. ‘unlike images’ represent two arbitrary images. Figure 2, shows two pairs, a pair of like images (Album-1 and Album-2) and a pair of unlike images (Poker and Wedding). The images were divided into three “lots” where the members of a “lot” share similar (but not necessarily the same) scenes. Overall, 277 pairs of images were processed by the length code algorithm.


Figure 2, Pairs of natural images

Figure 3, shows a summary of the results over the entire set of experiments (per lot) described in this section. The first 3 columns represent the average results for the three lots; the last column represents the overall average. It can be noticed that the SNR of “like” images is relatively high and is significantly better than the SNR of the ‘unlike images’. Hence, the SNR can be used to make an (automatic) decision whether to accept or reject the results of registration.

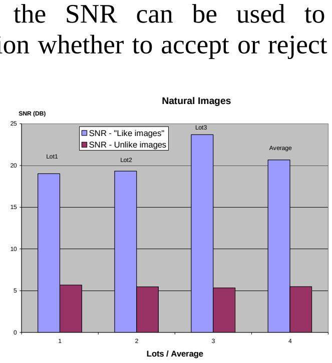


Figure 3, Natural Image Results

This also denotes that the registration is successful and the transformed image *T* (obtained from the image *I* after applying the inverse transformation $A^{-1}$ to the image *I)* is very similar to the reference image *R*. The results of this phase have also been evaluated manually. It was concluded that the registration error is very small. Moreover, the registration process can enable high quality change detection. It should be noted that the SNR obtained in the current phase is better than the SNR obtained with artificial transformations applied to natural images. We have concluded that the reason is due to the severity of the transformations parameters applied in the set of experiments with artificial transformations.

## Result Evaluation

Hundreds of experiments were performed. Due to the space limitations only a few of them have been presented. In general, when affine transformations were a composition of translation, rotation, and scaling, favorable objects matching and image registration results were obtained. Shearing and composite transformations are a bit more problematic. Overall, the DP algorithm has been shown to be successful for object matching.

Other problem areas identified include the tight coupling of object matching to segmentation, finding additional useful error measure between images after the computed transformations are applied, and guaranteeing the correct matching of objects as part of the DP process.

Finally, we have compared our registration with other registration algorithms reported in the literature and found that they are much better than published results (Brown 1992, Zitova 2003).

## Conclusions

This is the first research on using DSW in the length code domain for image registration. The results indicate that they merit further investigation including:

- It may be possible to use the matching function obtained when the DSW is applied to pairs of objects to estimate the global transformation applied to the entire image.
- Investigation of efficient ways to obtain more than three points with fewer than three matches.
- Investigation of the utility of different ways to obtain the length code (e.g., using an equi-angle signature), as well as other object shape features.
- Further investigation of the applied DSW constraints can improve the recognition rate of the DSW.
- The segmentation process is tightly bound to the DP matching algorithm. The impact of other segmentation methods on the DSW based registration algorithm can be investigated.
- Our experience shows that the building blocks of the algorithm are relatively efficient. Nevertheless, further investigation into computational complexity can better quantify the efficiency of the algorithms.
- Identifying additional registration quality criteria (in addition to SNR)

## References


Alajlan N. et al, *Shape Retrieval Using Triangle-Area Representation and Dynamic Space Warping*, Technical Report, University of Waterloo, 2006.

Brown L. G., “A survey of Image Registration Techniques” *ACM Computing Surveys*, 1992.

Baggs R., Lam T., and Tamir D.E., "Image Registration Using the Length Code Algorithm", *IEEE Southeastcon*, 1996.

Baggs R., *Object Matching Using the Dynamic Programming Algorithm*, PhD, Dissertation, Florida Institute of Technology 1997.

Bellman R. and Kalaba R., *Dynamic Programming and Modern Control Theory*, Academic Press 1965.

Coleman, G., Andrews, H., “Image Segmentation by Clustering”, *Proceedings of the IEEE*, 1979: 773-785.

Foley J. F., et al, *Computer Graphics: Principles and Practice*, Second edition, Addison Wesley, 1990.

Freeman H., “computer Processing of Line Drawing Images”, Computer Surveys 1974.

Haralick R. M., Shapiro G., *Computer and Robot Vision*, Vol. I and II, Addison Wesley 2002

Itakura F., “Minimum Prediction Residual Applied to Speech Recognition”, IEEE Transactions on ASSP, 1975.

Keogh E. et al, “LB_Keogh Supports Exact Indexing of Shapes under Rotation Invariance with Arbitrary Representations and Distance Measures”, *VLDB* 2006.

Linde, Y., Buzo, A. and Gray, R., “An Algorithm for Vector Quantizer Design,” I*EEE Transactions on Communications*, vol. COM-28, no 1, 1980.

Sahbi, H. et al, “Robust matching by dynamic space warping for accurate face recognition”, Proceedings. *Of the International Conference on Image Processing*, 2001.

Tamir D. E., Park C., and Yoo W., "The validity of pyramid K-means", *SPIE conference on Optics,* 2007.

Zitova B., and Flusser J., “Image registration methods: a survey”, Image and Vision Computing, 2003: 977-1000.