

# Multi-modal 3D Image Registration Based on Estimation of Non-rigid Deformation

Roberto Rosas-Romero, Oleg Starostenko, Jorge Rodríguez-Asomoza,  
and Vicente Alarcon-Aquino

Department of Computing, Electronics, and Mechatronics,  
Universidad de las Américas Puebla, Cholula, 72820, México  
{roberto.rosas, oleg.starostenko, jorge.rodriguez,  
vicente.alarcon}@udlap.mx

**Abstract.** This paper presents a novel approach for registration of 3D images based on optimal free-form rigid transformation. A proposal consists in semiautomatic image segmentation reconstructing 3D object surfaces in medical images. The proposed extraction technique employs gradients in sequences of 3D medical images to attract a deformable surface model by using imaging planes that correspond to multiple locations of feature points in space, instead of detecting contours on each imaging plane in isolation. Feature points are used as a reference before and after a deformation. An issue concerning this relation is difficult and deserves attention to develop a methodology to find the optimal number of points that gives the best estimates and does not sacrifice computational speed. After generating a representation for each of two 3D objects, we find the best similarity transformation that represents the object deformation between them. The proposed approach has been tested using different imaging modalities by morphing data from Histology sections to match MRI of carotid artery.

**Keywords:** 3D image matching, non-rigid deformation estimation, wavelet.

## 1 Introduction

Estimation of non-rigid deformation associated with objects in medical images can be used as assisting clinical tool to identify abnormal organ behavior. Medical image registration can be performed first by finding the deformation between a pair of images and then correcting the changes associated with such deformations to perform multi-modal image registration between medical data sets to integrate information from different modalities (ultrasound, X-ray, Magnetic Resonance Imaging (MRI), Histology, etc.), as well as register images taken at different times (temporal registration). So that, changes associated with disease evolution can be inferred. Since the resolution and distortion are different in every imaging modality and the tissue often changes in size and shape with time, estimation of deformation is a growing field.

Besides its medical application, image registration also has a variety of other applications, such as aerial image analysis, stereo vision, automated cartography, motion analysis, recovery of 3-D characteristics of a scene, and performing morphing of 3-D data sets for computer animation of visual effects [1-4]. One problem in the most efforts to describe organ deformations is that they require extensive human interaction. For instance, visual inspection of shape of the heart within a cardiac cycle is widely used to detect abnormalities [5]. In other cases, deformation is estimated by manually tracking the movement of predefined landmark points of organ borders. Techniques that create markers on images have the drawbacks of being invasive and the fact that the more tags there are, the poorer the background image signal to noise is. Thus, we propose a non-invasive technique that reduces the amount of human intervention.

Usually, image registration is accomplished first by systematically reconstructing the surfaces and feature point sets of two 3D objects extracted from two sets of images. Therefore, for each set there is a representation of the object that consists in its surface and a set of feature points. Feature points to be extracted are those that can be used as a reference before and after a deformation. After generating a representation for each of the two 3D objects, we find the best similarity transformation that represents the object deformation between them. Well-known methods based on the registration of 3D curves are efficient enough but these methods are not useful when registration of 3D surfaces is required [7], [8]. Other efforts try to model non-rigid a deformation using successive transformation such as twisting, banding, tapering and the inconvenience of these approaches is that a non-rigid deformation might require a description, which is not provided by a combination of simple deformation models [9]. Registration of 3-D medical images under non-rigid deformation using physical properties of the objects has been widely studied, however, one problem of these techniques is that physical properties must be obtained for each specific application and they might be difficult to obtain or not available [10], [11].

## 2 3D Object Surface Extraction Based on Active Contour Models

Extraction of the surface of an object from an image set consists in reconstructing its shape from points collected from its physical surface. There is a set of 3D images that describes an object that is used as a reference. A set of imaging planes is obtained by scanning the object in parallel slices and the intersection of each imaging plane with the object gives a contour. Tracing of contour points on parallel imaging planes and joining them generates a 3D surface. We use active contour models to extract contour points from images [12].

Consider the problem of detecting the borders of a 3D reference object. If there are  $m$  planes per 3D object, and  $n$  points per plane; then there are  $N = mn$  contour points to be searched. By using the Original Snake Active Contour Model, a single imaging plane is used to detect  $n$  contour points on that specific plane, and this process is repeated for each of  $m$  imaging planes in the object. Instead of attempting to perform contour detection for each imaging plane in isolation, we directly approach it as a 3D problem; so that the  $mn$  contour points corresponding to the object surface are

detected at once from multiple imaging planes. A snake is a deformable curve, which approaches contours on images by optimizing the placement of the snake points that form the curve. In the 3D model, each snake point

$$\mathbf{v}(r,s) = [x(r,s), y(r,s), z(r,s)]^T \quad (1)$$

is a function of two parameters  $r$  (spatial index),  $s$  (imaging plane index); so that the 3D snake function to be optimized is defined as

$$f(\mathbf{v}) = \alpha_1 \| \mathbf{v}_r \| + \alpha_2 \| \mathbf{v}_s \| + \beta_1 \| \mathbf{v}_{rr} \| + \beta_2 \| \mathbf{v}_{ss} \| + \beta_4 \| \mathbf{v}_{rs} \| + E \quad (2)$$

where  $\{\alpha_i\}$  are constants imposing a tension constraint, and  $\{\beta_i\}$  are constants imposing a bending constraint and  $E$  is some sort of image gradient function. Since the snake points stick to their plane, they are functions of the  $x$  and  $y$  coordinates only. The minimization of the snake energy function yields two Euler equations which can be iteratively solved to find a local minimum of the snake energy function,

$$\begin{aligned} (\mathbf{D} + \gamma \mathbf{I}) \mathbf{x}_i &= \gamma \mathbf{x}_{i-1} - \mathbf{f}_x(\mathbf{x}_{i-1}, \mathbf{y}_{i-1}) \\ (\mathbf{D} + \gamma \mathbf{I}) \mathbf{y}_i &= \gamma \mathbf{y}_{i-1} - \mathbf{f}_y(\mathbf{x}_{i-1}, \mathbf{y}_{i-1}) \end{aligned} \quad (3)$$

where  $\mathbf{D}$  is an  $N \times N$  penta-diagonal matrix; the vectors  $\mathbf{x}_i$  and  $\mathbf{y}_i$  are the coordinates of the snake points at the  $i$ th iteration; the vectors  $\mathbf{f}_x(\mathbf{x}_i, \mathbf{y}_i)$  and  $\mathbf{f}_y(\mathbf{x}_i, \mathbf{y}_i)$  are the image forces of the partial derivatives of the external energy for the snake points; and  $\gamma$  is the step size control parameter. Since the matrix  $\mathbf{D}$  is quite big, it is impractical to invert it directly, it is assumed that the change between  $\mathbf{x}_i$  and  $\mathbf{x}_{i-1}$  is small enough and LU decomposition is used for each plane.

### 3 Extraction of 3D Object Feature Points from Sets of Images

For feature extraction, similar regions of the object from two different sets of images are manually identified with the help of a radiologist. Then two feature points, from the sets of images that contain each of the identified regions, are extracted; so that feature extraction and feature correspondence establishment are accomplished simultaneously. Each selected feature point is an edge point whose edge response is maximum within the identified region. This edge-based approach for extraction of pairs of corresponding feature points from 3D regions applies the Wavelet Transform [13].

Let's consider two objects  $O$  and  $O'$ , which are related with a non-rigid deformation. To find a pair of correctly-matched feature-points from these two objects, we must first manually identify a 3D region from the set of imaging planes  $\{I_1, I_2, \dots, I_m\}$  that describes  $O$ , and also a similar region from the set  $\{I'_1, I'_2, \dots, I'_n\}$  corresponding to  $O'$ . A region of interest in  $O$  is defined as a 3D discrete function  $f(x, y, z)$  that gives a gray level value. At  $z = z_r$ ,  $f(x, y, z_r)$  corresponds to a rectangular window within one particular imaging plane from the set of images  $\{I_i\}$ . Similarly, a region of interest in  $O'$  is a discrete function  $g(x, y, z)$ , where  $g(x, y, z_s)$  corresponds to a rectangular window on one imaging plane from  $\{I'_j\}$ . Basically, these discrete functions are generated by extracting sections from images and stacking them. The condition used to identify regions of interest is that they must contain structures that are common to both objects  $O$  and  $O'$ . These structures correspond to sharp variations, which are generally located at boundaries, edges or corners. Once  $f(x, y, z)$

and  $g(x, y, z)$  are established; one feature point  $P(x, y, z)$  is automatically extracted from  $f(x, y, z)$  and a corresponding point  $Q$  is obtained from  $g(x, y, z)$ . The pair  $(P, Q)$  is called a correctly-matched feature-point pair. Wavelet transform for multiresolution local analysis is applied to extract these points.

Let  $S(x, y, z)$  be a 3D smoothing function. We call a smoothing function any function  $S(x, y, z)$  equal to a Gaussian. Three wavelets,  $\psi^1(x, y, z)$ ,  $\psi^2(x, y, z)$  and  $\psi^3(x, y, z)$  are the partial derivates of the smoothing function  $S(x, y, z)$  in the  $x$ ,  $y$  and  $z$  directions, respectively, where:

$$\begin{aligned}\psi^1(x, y, z) &= \frac{\partial S(x, y, z)}{\partial x}, \quad \psi^2(x, y, z) = \frac{\partial S(x, y, z)}{\partial y}, \\ \psi^3(x, y, z) &= \frac{\partial S(x, y, z)}{\partial z}\end{aligned}\quad (4)$$

Dilating these functions by a scaling factor  $2^j$ ,

$$\begin{aligned}\psi_j^1(x, y, z) &= \frac{1}{8^j} \psi^1\left(\frac{x}{2^j}, \frac{y}{2^j}, \frac{z}{2^j}\right), \quad \psi_j^2(x, y, z) = \frac{1}{8^j} \psi^2\left(\frac{x}{2^j}, \frac{y}{2^j}, \frac{z}{2^j}\right), \\ \psi_j^3(x, y, z) &= \frac{1}{8^j} \psi^3\left(\frac{x}{2^j}, \frac{y}{2^j}, \frac{z}{2^j}\right)\end{aligned}\quad (5)$$

At each scale,  $2^j$ , the 3D wavelet transform of a function  $f(x, y, z)$  can be decomposed into three directions as,

$$\begin{aligned}W_j^1 f(x, y, z) &= f(x, y, z) * \psi_j^1(x, y, z); \\ W_j^2 f(x, y, z) &= f(x, y, z) * \psi_j^2(x, y, z); \\ W_j^3 f(x, y, z) &= f(x, y, z) * \psi_j^3(x, y, z).\end{aligned}\quad (6)$$

These three components are equivalent to the gradients of  $f(x, y, z)$  smoothed by  $S(x, y, z)$  at scale  $2^j$  in the  $x$ ,  $y$  and  $z$  directions. The local extreme of  $W_i^1 f(x, y, z)$ ,  $W_i^2 f(x, y, z)$  and  $W_i^3 f(x, y, z)$  corresponds to the inflection points of surface  $f(x, y, z) * S(x, y, z)$  along the  $x$ ,  $y$  and  $z$  directions respectively. The direction of a gradient vector at a point  $(x_0, y_0, z_0)$  indicates the direction in the space  $(x, y, z)$  along which the directional derivate of  $f(x, y, z)$  has the largest absolute value. Three-dimensional edges are defined as points  $(x_0, y_0, z_0)$ , where the modulus of gradient vector is maximum. Hence, 3D edge points can be located from the three components of  $W_i^1 f(x, y, z)$ ,  $W_i^2 f(x, y, z)$  and  $W_i^3 f(x, y, z)$  of the wavelet transform. At a specific scale  $2^j$ , the modulus of the gradient vector of  $f(x, y, z)$  can be calculated as,

$$M_j f(x, y, z) = \sqrt{\left|W_j^1 f(x, y, z)\right|^2 + \left|W_j^2 f(x, y, z)\right|^2 + \left|W_j^3 f(x, y, z)\right|^2} \quad (7)$$

If the local maxima of  $M_j f(x, y, z)$  are located, then all the 3D edge points of  $f(x, y, z)$  at scale  $2^j$  can be detected.

In general, noise is the main cause of false detection of edge points. In order to suppress the effect of noise, a criterion called edge correlation is introduced,

$$R_n(j, x, y, z) = \prod_{i=0}^{n-1} M_i + i f(x, y, z) \quad (8)$$

where  $n$  is a positive number indicating the number of scales involved in the multiplication, and  $j$  represent the initial scale for edge correlation. This process detects edge points, whose edge responses are the strongest within a local area. Two conditions are adopted to judge whether a point  $(x_o, y_o, z_o)$  is a feature point or not:

*Condition 1.*  $(x_o, y_o, z_o)$  must be a 3D edge point of the function  $f(x, y, z)$ . This means that  $(x_o, y_o, z_o)$  is a local maxima of  $M_j f(x, y, z)$ .

*Condition 2.*  $M_j f(x_o, y_o, z_o) = \max \{M_j f(x, y, z) \mid (x, y, z) \in N_p\}$ , where  $N_p$  is the region represented by  $f(x, y, z)$ .

## 4 Estimation of Non-rigid Deformation

Once a set of surface points  $S_1$  and a set of feature points  $FP_1$  in the object are established for each set of images, we need to find the transformation function  $T$  that matches the sets  $S_1$  and  $FP_1$  to the sets  $S_2$  and  $FP_2$ ,  $T(\{S_1, FP_1\}, p) \approx \{S_2, FP_2\}$ , where  $p$  are the transformation parameters that have to be found. The search of the deformation parameters is an optimization process that minimizes the *differences* between two sets of points (Levenberg-Marquardt Least Squares Minimization) [14]. During this optimization process, deformations are systematically applied to  $S_1$  and  $FP_1$ , by adjusting the set of parameters  $p$ , until the corresponding sets of transformed points  $T(\{S_1, FP_1\}, p)$  get as close as possible to the sets  $\{S_2, FP_2\}$ , until the distance  $d()$  between both sets is minimized. Thus, estimating deformation can be referred to as the minimization of the cost function,  $C(p) = d(\{S_2, FP_2\}, T(\{S_1, FP_1\}))$ .

For a similarity metric, the distance function establishes a parametric representation of the 3D object surface, using imaging planes at the first time frame, and enables measurement of 3D deformation during object movement within a time sequence. After a distance function is constructed for the initial 3D shape, the tracked surface based on the snake is fed to the distance function to perform deformation estimation. The model used for a 3D distance function is based on the 2D *chamfer distance model* [15]. Tri-linear interpolation is used to transform the resulting discrete distance map into a continuous distance function. In this interpolation process, the distance from any point  $r$  to  $S_2$  is computed by finding the eight grid points which form the voxel that contains  $r$  and then interpolate the distance value  $d(r, S_2)$  contributed by the distance values  $d_{ijk}$  at the eight vertices of the voxel.

Assume there is a total of  $N$  points on the sample surface  $\{q_i \mid i = 1, 2, \dots, N\}$ , and the corresponding points after transformation are  $\{r_i = T(q_i, p) \mid i = 1, 2, \dots, N\}$ . It is defined that  $d_i$  is the distance between  $r_i$  and the nearest point on the reference surface. There is a total of  $N$  distance terms from each transformed point to the reference surface, so that the cost function can be formulated as

$$C(p) = \sum_{i=1}^N d_i(p)^2 \quad (9)$$

*Free Form Deformation FFD Models* are used to describe non-rigid deformations. These models describe the deformation of an object in terms of the deformation of the space that contains the object. The set of parameters  $\mathbf{p}$  for the transformation function  $T$  consists of a set of deformation vectors  $\{\mathbf{v}_{ijk}\}$  located in a 3D discrete grid.

The mathematical model for the function that represents the free-form deformation of an arbitrary point corresponds to a mapping from the source coordinates  $(x, y, z)$  to the transformed coordinates  $(x', y', z')$ .

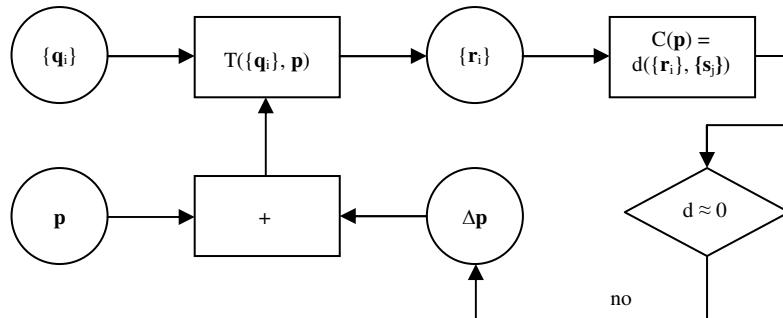
$$[x', y', z']^T = [x, y, z]^T + \mathbf{v}(x, y, z) \quad (10)$$

where the displacement function  $\mathbf{v}(x, y, z)$  is a linear combination of interpolating functions.

$$\mathbf{v}(x, y, z) = \sum_i \sum_j \sum_k \psi_{ijk}(x, y, z) \mathbf{v}_{ijk} \quad (11)$$

with the set of interpolating functions  $\{\psi_{ijk}(x, y, z) = \psi_i(x)\psi_j(y)\psi_k(z); i = 0, \dots; j = 0, \dots; k = 0, \dots\}$  generated from a first-order spline function  $\psi$  by simple translation. As the indexes  $(i, j, k)$  change, the location of the function  $\psi_{ijk}$  moves along a 3D grid. For each point grid, there is only one function with non-zero value.

A general block diagram for estimation of deformations is shown in Fig. 1.

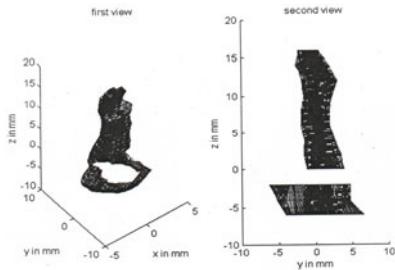


**Fig. 1.** Block diagram for the optimization process to estimate non-rigid deformations

## 5 Experiments and Results

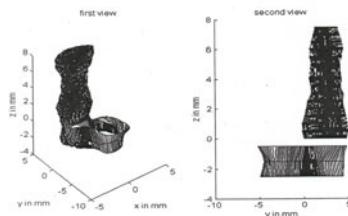
Experiments were conducted using real medical data. The reference objects for registration were the lumen of a carotid artery, and were conducted to register MRI and Histology data sets. The most common distortion between data from MRI and data from Histology is the shrinkage of the tissue during the histological process. Thus, to perform registration of an object extracted from these modalities, non-rigid deformation between them is estimated and then the object from Histology is morphed back to match the one from MRI.

During these experiments, sets of 16 MRI imaging planes were obtained from a section of the carotid artery, with each imaging plane represented by a *512 x 512 pixel* matrix over a field of view of *90 mm x 90 mm*, with a distance of *2 mm* between adjacent planes. The histological section of the same lumen was digitized to generate a set of 36 imaging planes, using a matrix of *480 x 512 pixels* over a field of view of *170 mm x 180 mm*, with variable distances between adjacent slices. Fig. 2 shows different views of the reconstruction of the lumen surface from the set of MRI images.



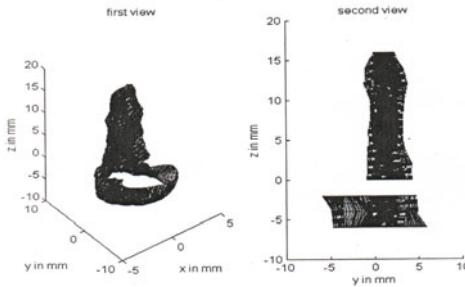
**Fig. 2.** Surface reconstruction from MRI data of carotid artery

Fig. 3 shows the corresponding reconstruction from Histology. Images from both modalities were used to extract feature points from regions of interest, and the criterion used to identify these features was that they had to contain structures common to both modalities, with such structures corresponding to sharp variations generally located at boundaries, edges or corners. Therefore, for each region of interest in the Histology set, there is a similar region selected from the MRI set.



**Fig. 3.** Carotid artery surface reconstruction from Histology data

Surface and feature points from Histology data are matched to those from MRI by estimating non-rigid deformation between both modalities. These estimates took less than 5 minutes on a PC machine and required initialization of the Levenberg-Marquardt algorithm by setting the deformation parameters to zero. After performing 10 iterations of rigid matching followed by 40 iterations of non-rigid matching, the registered data sets appear in Fig. 4.



**Fig. 4.** Matching of histology data to MRI data

To measure the error for this matching, the distance between the set of feature points extracted from the MRI images and the set of feature points from Histology images after matching (in different combinations: Object1 from MRI - Object 2 from Histology, Object1 from MRI - Object 2 from Histology MRI, and Object1 from Histology- Object 2 from Histology) is estimated by computing the root mean square error between both sets. The average number of feature points from object 1 and objects 2 is 15 for each one. The estimated absolute errors for these experiments were 3.23, 2.1, and 1.23 mm respectively. Table 1 shows the error corresponding to the matching of two objects after performing 68 experiments for multi-modal image registration.

**Table 1.** Average estimated error for 68 different experiments on registration

Modalities of matching images	Number of experiments	Average error of two object matching	Average related error of object matching
MRI - Histogram	22	3.32 mm	3.68 %
MRI – MRI	34	1.23 mm	1.36 %
Histogram - Histogram	12	2.65 mm	2.94%

## 6 Conclusions

This paper presents a new technique for multi-modal image registration based on the estimation of non-rigid deformation in the three dimensional space. The specific case under study consists in data registering sets of data from different imaging modalities by the morphing data from histology sections to match MRI. The effectiveness and accuracy of the deformation estimates depend on the number of surface points and the number of feature points extracted from sets of medical images. An issue concerning this relation is difficult and deserves attention to develop a methodology to find the optimal number of points that gives the best estimates and does not sacrifice speed of computation. In order to obtain a set of correctly-matched feature-point pairs, our approach requires selections of similar regions of interest between two imaging modalities. Consequently, it also requires manual establishment of correspondence between two sets of features. In order to avoid manual selection of regions of interest,

we have suggested the automatic extraction of feature points from the whole region described in a sequence of images. The obtained results show satisfactory functionality of the proposal, particularly, the related error of image matching is on about 3% for different modalities of image sets with dimensions about 90x90 mm. The disadvantage consists in used simple conditions to judge the selection of feature points. One way to automatically establishing correspondence between two sets of feature points is the use of combinatorial search. This will require the development of a measurement of the similarity for two features points that must overcome the differences between two target images.

**Acknowledgments.** This research is sponsored by Mexican National Council of Science and Technology, CONACyT #109115 and #109417.

## References

1. Mironenko, A., Song, X.B.: Image registration by minimization of residual complexity. In: IEEE Computer Soc. Conf. on Computer Vision and Pat. Recog. USA, pp. 49–56 (2009)
2. Xai, M., Liu, B.: Image Registration by Super-Curves. IEEE Transactions on Image Processing 13(5) (2004)
3. Zhu, Z., Hanson, A.R., Riseman, E.M.: Generalized Parallel-Perspective Stereo Mosaics from Airborne Video. IEEE Trans. on Pattern Analysis and Machine Intel. 26(2) (2004)
4. Adiga, U., Malladi, R., Gonzalez, R., Ortiz, C.: High-Thoughput Analysis of Multispectral Images of Breast Cancer Tissue. IEEE Transactions on Image Processing 15(8) (2006)
5. Moore, C.C., et al.: Three-dimensional Systolic Strain Patterns in the Normal Human Left Ventricle: Characterization with Tagged MR Imaging. Radiology 214, 453–466 (2000)
6. Yau, H.T., Tsou, L.S., Tseng, H.M.: Automatic Registration Using Virtual Polar Ball. Computer-Aided Design & Applications 4(1-4), 427–436 (2007)
7. Pouderoux, J.: Global Contour Lines Reconstruction in Topographic Maps (2007)
8. Sumengen, B., Manjunath, B.S.: Graph Partitioning Active Contours (GPAC) for Image Segmentation. IEEE Trans. on Pattern Analysis and Machine Intelligence 28(4) (2006)
9. Lazaridis, G., Petrou, M.: Image Registration Using the Walsh Transform. IEEE Transactions on Image Processing 15(8) (2006)
10. Zayer, R., Rossl, C., Karmi, Z., Seidel, H.: Harmonic Guidance for Surface Deformation Journal: Computer Graphics Forum, vol. 24(3), pp. 601–609 (2005)
11. Kempeeneers, P., et al.: Generic Wavelet-Based Hyperspectral Classification Applied to Vegetation Stress Detection. IEEE Trans. on Geoscience and Remote Sensing 43(3) (2005)
12. Kuman, R.: Snakes, Active Contour Models: Implements snakes or active contour models for image segmentation, Matmal (2010)
13. Alarcón-Aquino, V., Starostenko, O., et al.: Initialisation and Training Procedures for Wavelet Networks Applied to Chaotic Time Series. J. of Eng. Intelligent Systems 18(1), 1–9 (2010)
14. Gill, P.E., Murray, W.: Practical Optimization. Academic Press, New York (1981)
15. Borgefors, G.: Digital Transformations in Digital Images. Computer Vision, Graphics and Image Processing 34 (1986)