INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN BIOMEDICAL ENGINEERING
Int. J. Numer. Meth. Biomed. Engng. 2015; **00**:1–63
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/cnm

Energy Minimization in Medical Image Analysis: Methodologies & Applications

Feng Zhao and Xianghua Xie*

Department of Computer Science, Swansea University, Swansea, SA2 8PP, United Kingdom

SUMMARY

Energy minimization is of particular interest in medical image analysis. In the past two decades, a variety of optimization schemes have been developed. In this paper, we present a comprehensive survey of the state-of-the-art optimization approaches. These algorithms are mainly classified into two categories: continuous method and discrete method. The former includes Newton-Raphson method, gradient descent method, conjugate gradient method, proximal gradient method, coordinate descent method, and genetic algorithm-based method, while the latter covers graph cuts method, belief propagation method, tree-reweighted message passing method, linear programming method, maximum margin learning method, simulated annealing method, and iterated conditional modes method. We also discuss the minimal surface method, primal-dual method, and the multi-objective optimization method. In addition, we review several comparative studies that evaluate the performance of different minimization techniques in terms of accuracy, efficiency, or complexity. These optimization techniques are widely used in many medical applications, e.g., image segmentation, registration, reconstruction, motion tracking, and compressed sensing. We thus give an overview on those applications as well. Copyright © 2015 John Wiley & Sons, Ltd.

Received . . .

KEY WORDS: energy minimization; optimization; graph cuts; deformable models; medical image segmentation; registration

1. INTRODUCTION

Medical imaging is becoming more and more prominent for disease diagnosis as well as treatment planning. However, the medical image data collected using different sorts of imaging modalities, e.g., magnetic resonance imaging (MRI), computed tomography (CT), ultrasound, along with positron-emission tomography (PET) are mostly of high ambiguity and complexity, due to the inherent limitations of the imaging procedure like artifact, inhomogeneity, various noises, restricted resolution, and inadequate contrast. To locate the target objects along with their boundaries, image

*Correspondence to: Department of Computer Science, Swansea University, Swansea, SA2 8PP, United Kingdom. E-mail: x.xie@swansea.ac.uk

Contract/grant sponsor: National Institute for Social Care and Health Research (NISCHR)

segmentation techniques are popularly applied, which can represent the volumetric image data in a more meaningful way, thus makes them simpler to be analyzed. The outlining of these closely connected anatomical structures in medical image data is critical to various medical applications such as image enhancement, image reconstruction, disease diagnosis, surgical planning, and 3D image visualization.

Extracting anatomical structures and subsequently rebuilding a concise geometric description of them is a tough task because of the variability and complexity of the object shapes along with the variation in image quality. Particularly, medical images usually contain different sorts of noises and sampling artifacts that may create disconnected and indistinct boundaries. Conventional low-level image processing techniques based on local intensities (e.g., edge detection, thresholding) generally yield impracticable object boundaries. Among many other approaches, energy minimization is particularly popular in medical image segmentation, where a function is first formulated to evaluate the goodness of a segmentation for a certain image, and the resulting segmentation of this image is given by minimizing such a function.

Deformable models are one of the most popular and promising methods for segmentation employing energy minimization, which combine geometry, physics, and approximation theory to provide powerful approaches to medical image analysis. The earliest approach is the active contour models (also known as snakes) suggested by Kass et al. [1] for the extraction of whole object boundaries in images, where a snake is a continuously deformable curve restricted by interior smoothness constraints and exterior image constraints. Since snakes paved the way, there have been numerous refinements making every effort to increase the accuracy, robustness, and efficiency. These models are successful in the segmentation, matching, along with tracking of the anatomies in medical images, making full use of constraints obtained from the medical data, and the incorporation of a priori knowledge of these structures' location, size as well as shape. They have the ability of representing the complicated shape together with large shape variations in different anatomical structures. In [2, 3], McInerney and Terzopoulos presented an overview on earlier deformable models popularly applied in the analysis of medical images. Later, Xu et al. [4] explained the definition of geometric and parametric deformable models for the recovering of shape boundaries. They also gave a survey on a few essential generalized deformable models, in which the prior knowledge/information specified by applications together with global properties of shapes are utilized to achieve high robustness and accuracy. Although mostly dissimilar in nature, the applications of these approaches to medical image segmentation are often constructed on five predominant components: energy/cost function, segmentation characterization, image description, training images, together with minimizer. The authors in [5] provided an in-depth discussion of the main work in each component. Recently, Zhao and Xie [6] presented an overview on interactive segmentation techniques for medical images, including multiple energy minimization-based methods.

Energy minimization techniques are also popularly employed in many other medical applications like image registration, reconstruction, motion tracking, and compressed sensing. In the past decades, a number of new approaches to energy minimization have been proposed in the literature. Thus, it is worth giving a systematic review on these vigorously investigated approaches. In this survey, we classify these state-of-the-art energy minimization algorithms into two main categories: continuous method and discrete method. The former largely includes Newton-Raphson method,

DOI: 10.1002/cnm

gradient descent method, conjugate gradient method, proximal gradient method, coordinate descent method, and genetic algorithm-based method, while the latter mostly covers graph cuts method, tree-reweighted message passing method, belief propagation method, linear programming method, maximum margin learning method, simulated annealing method, and iterated conditional modes method. We also discuss the minimal surface method, primal-dual method, and the multi-objective optimization method. Additionally, we investigate several comparative studies that evaluate the performance of different minimization techniques in terms of accuracy, efficiency, or complexity.

The remainder of this survey is structured as follows. In Section 2, we give the definitions and foundations of energy minimization. The variety of energy minimization techniques is thoroughly described in Section 3. Following that, we discuss the applications of these approaches in medical image analysis in Section 4. The conclusions are highlighted in Section 5.

2. ENERGY FUNCTION

In energy minimization, the first step is to attain the form of the energy/cost function (objective function), which can be formulated from different aspects and then be minimized using appropriate optimization techniques. In this section, we focus on several representative energy formulations.

2.1. Variational formulation

This class of energy function is generally defined as a weighted sum of both internal energy and external energy. Minimization of the overall energy produces the internal and external forces, where the former keeps the curve/contour smooth without large bending, while the latter draws the curve/contour towards the desired image features such as object boundaries. Assume that $I(x, y)$ denotes a gray-scale image, $\boldsymbol{c}(k) = (x(k), y(k))$ describes a contour, where x, y are functions of the coordinates, $k \in [0, 1]$ stands for the parametric domain, we can write its energy functional as,

$$E(\boldsymbol{c}(k)) = E_{int}(\boldsymbol{c}(k)) + E_{ext}(\boldsymbol{c}(k)). \tag{1}$$

The first part in Eq. (1): $E_{int}(\cdot)$, denotes the internal energy functional, defined as,

$$\begin{aligned} E_{int}(\boldsymbol{c}(k)) &= \frac{1}{2}\int_0^1 \left(\omega_1(k)|\boldsymbol{c}_k(k)|^2 + \omega_2(k)|\boldsymbol{c}_{kk}(k)|^2\right) dk, \\ &= \frac{1}{2}\int_0^1 \left(\omega_1(k)\left|\frac{\partial \boldsymbol{c}}{\partial k}\right|^2 + \omega_2(k)\left|\frac{\partial^2 \boldsymbol{c}}{\partial k^2}\right|^2\right) dk, \end{aligned} \tag{2}$$

where the first- and second-order derivatives inhibit stretching and bending individually, but make the deformation of the contour act similar to an elastic string and a rigid rod separately, while ω_1 and ω_2 restrict the contour's 'tension' and 'rigidity' respectively.

The second part in Eq. (1): $E_{ext}(\cdot)$, denotes the external energy functional, defined as,

$$E_{ext}(\boldsymbol{c}(k)) = \int_0^1 Q(\boldsymbol{c}(k))dk, \tag{3}$$

Prepared using cnmauth.cls
DOI: 10.1002/cnm

where $Q(\cdot)$ is a scalar potential function derived from the images and it takes smaller values at the desired image features (e.g., object boundaries, intensity edges). Considering the gray-scale image I as a function with respect to continuous coordinate variables (x, y), a representative potential function pulling a deformable contour toward intensity edges in this image can be designed as,

$$Q(x,y) = -\omega_c \left|\nabla[G_\sigma(x,y) * I(x,y)]\right|^2, \tag{4}$$

where ω_c is a positive weighting parameter controlling the potential function's magnitude, ∇ denotes the gradient operator, $G_\sigma(\cdot)$ stands for a 2D Gaussian smoothing function with standard deviation σ controlling the spatial extent of the local optima of $Q(\cdot)$, and $*$ represents the 2D convolution operator.

As stated by the calculus of variations lemma, contour $\boldsymbol{c}$ that minimizes the energy functional $E(\cdot)$ should adequately comply with the Euler-Lagrange equation [1, 7],

$$\frac{\partial}{\partial k}\left(\omega_1 \frac{\partial \boldsymbol{c}}{\partial k}\right) - \frac{\partial^2}{\partial k^2}\left(\omega_2 \frac{\partial^2 \boldsymbol{c}}{\partial k^2}\right) - \nabla Q(\boldsymbol{c}(k)) = 0, \tag{5}$$

where the first two elements of this equation are called the internal forces that discourages stretching and bending respectively, and the third component is the external force which draws the contour towards the desired object boundaries. This partial differential equation (PDE) conveys the state in which internal and external forces are balanced, in case that the contour stops moving at equilibria.

In medical image analysis, sometimes it is more convenient to formulate a dynamic model, as the majority of anatomical structures in medical data are deformable and continuously undertake non-rigid motions. By treating the contour $\boldsymbol{c}$ as a time-varying one: $\boldsymbol{c}(k,t) = (x(k,t), y(k,t))$, in accordance with Newton's second law, the contour $\boldsymbol{c}(k,t)$ dynamics should meet the following equation,

$$F_{inert}(\boldsymbol{c}) = F_{damp}(\boldsymbol{c}) + F_{int}(\boldsymbol{c}) + F_{ext}(\boldsymbol{c}). \tag{6}$$

Equivalently,

$$\mu \frac{\partial^2 \boldsymbol{c}}{\partial t^2} = -\gamma \frac{\partial \boldsymbol{c}}{\partial t} + \frac{\partial}{\partial k}\left(\omega_1 \frac{\partial \boldsymbol{c}}{\partial k}\right) - \frac{\partial^2}{\partial k^2}\left(\omega_2 \frac{\partial^2 \boldsymbol{c}}{\partial k^2}\right) - \nabla Q(\boldsymbol{c}(k,t)), \tag{7}$$

where μ is the mass coefficient, γ indicates the damping coefficient. The left-hand side of this PDE is the inertial force, the first component on the right-hand side denotes the damping force, and similar to Eq. (5), the second and third parts characterize the internal forces of stretching and bending respectively, while the last one stands for the external force. Equilibrium will be obtained if the internal and external forces are balanced, and simultaneously the contour stops moving, that is, $\frac{\partial \boldsymbol{c}}{\partial t} = 0$ and $\frac{\partial^2 \boldsymbol{c}}{\partial t^2} = 0$.

In addition, many approaches have been developed by designing novel energy terms including both external and internal terms to address a particular class of images or alleviate a particular problem with the original terms. For example, Caselles et al. [8] proposed an energy functional defined as follows,

$$E_{GAC}(\boldsymbol{c}(k)) = \int_0^{L_{\boldsymbol{c}}} e(|\nabla I(\boldsymbol{c}(k))|)dk, \tag{8}$$

DOI: 10.1002/cnm

where I and $\boldsymbol{c}$ are the original image and the evolving contour, $L_{\boldsymbol{c}}$ represents the Euclidean length of contour $\boldsymbol{c}$, while e stands for a function indicating edges, which disappears at boundaries of the objects in image I. The function e embraces information relating to the object boundaries, and an example of this function could be $e(|\nabla I|) = 1/(1+\beta|\nabla I|^2)$ where β means a constant with positive value. A rapid global minimization algorithm was presented in [9], where a convex energy functional is formulated on the basis of the weighted total variation (TV) norm applying the primal-dual method [10, 11],

$$E_{TV_e}(v,\lambda) = \underbrace{\int_\Omega e(x)|\nabla v|dx}_{TV_e(v)} + \lambda \int_\Omega |v-g|dx, \tag{9}$$

where Ω denotes the image domain, $e(x)$ is the weighting function indicating the connection between the two energy functionals $E_{GAC}(\cdot)$ and $E_{TV_e}(\cdot)$, while v and g stand for the characteristic function and observed image respectively, $\lambda \geqslant 0$ marks a constraint parameter. In case that these functions and parameters satisfy specified conditions [6], it has been proved that $E_{GAC}(\cdot)$ and $E_{TV_e}(\cdot)$ are identical [9].

2.2. *Combinatorial formulation*

In graph-based works, the energy functional is often called cost function, which is composed of boundary information and regional information. It can be expressed as sums across the vertices along with their neighborhoods that vary as a function of how every vertex is assigned with a label. To produce a solution which is globally optimal, the cost function needs to be formulated appropriately.

Usually, an undirected graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ is formulated in terms of a group of nodes/vertices ($\mathcal{V}$) and undirected edges ($\mathcal{E}$) connecting the node pairs (v_p, v_q) [12, 13]. Every undirected edge ($e \in \mathcal{E}$) is associated with a weight/cost ($\omega_e \geqslant 0$). A graph cut $\mathcal{G}(\mathcal{C}) = \langle \mathcal{V}, \mathcal{E} \backslash \mathcal{C} \rangle$ ($\mathcal{C} \subset \mathcal{E}$) means a separation which splits edges into two terminals representing the object s and background t, respectively. The cost of a single cut is computed by adding all of the non-negative weights of those edges that it splits, that is, $|\mathcal{C}| = \sum_{e \in \mathcal{C}} \omega_e$.

In medical image analysis like segmentation, the pixels/voxels and the relationship of each pair of them are depicted by the graph nodes and edges, respectively. A graph cut partitioning those nodes indicates a potential image/volume segmentation, and the one with minimum cost yields an optimal segmentation with respect to image properties assigned into those non-negative edge weights. For simplicity, assume that M stands for the total number of pixels/voxels in an image/volume, $\mathcal{N}$ represents all the pairs of neighboring pixels/voxels under a 8-neighborhood (2D) or 26-neighborhood (3D) system, and $C = (C_{v_1}, \cdots, C_{v_p}, \cdots, C_{v_M})$ is a binary vector whose component C_{v_p} specifies the label ($\mathcal{O}$: 'object' or $\mathcal{B}$: 'background') assigned to a pixel/voxel v_p, the soft constraints imposed on the region and boundary properties of C could be delineated using the following cost function,

$$\begin{aligned} E(C) &= \lambda R(C) + B(C), \\ &= \lambda \sum_{v_p \in \mathcal{V}} R_{v_p}(C_{v_p}) + \sum_{\{v_p, v_q\} \in \mathcal{N}} B_{\{v_p, v_q\}} \delta(C_{v_p}, C_{v_q}), \end{aligned} \tag{10}$$

Prepared using cnmauth.cls
DOI: 10.1002/cnm

where

$$\delta(C_{v_p}, C_{v_q}) = \begin{cases} 1 & \text{if } C_{v_p} \neq C_{v_q} \\ 0 & \text{if } C_{v_p} = C_{v_q} \end{cases},$$

and $\lambda \geqslant 0$ is a parameter controlling the relative significance of the two terms: $R(C)$ and $B(C)$ corresponding to the region and boundary properties, respectively. The region term (also known as label cost) $R(C)$ implies the individual penalty for allocating a pixel/voxel v_p to label $\mathcal{O}$ or $\mathcal{B}$. For instance, based on the intensities of pixels or voxels labeled as $\mathcal{O}$ and $\mathcal{B}$ seed points by a user (known as hard constraints), the intensity distribution histograms for the object and background, denoted by $Pr(I|\mathcal{O})$ and $Pr(I|\mathcal{B})$, can be computed and then used for the calculation of the regional penalties,

$$\begin{aligned} R_{v_p}(\mathcal{O}) &= -\ln Pr(I_{v_p}|\mathcal{O}), \\ R_{v_p}(\mathcal{B}) &= -\ln Pr(I_{v_p}|\mathcal{B}). \end{aligned} \tag{11}$$

In the boundary term (also known as separation cost) $B(C)$, $B_{\{v_p,v_q\}} \geqslant 0$ is a coefficient reflecting the penalty for the discontinuity between pixels/voxels v_p and v_q. The coefficient can be calculated according to criteria such as local intensity gradient and gradient direction. Its value is big in case that v_p and v_q have similar properties, and near zero if the pixel/voxel pair is rather distinct. As an example, the boundary penalties can be set using the following function,

$$B_{\{v_p,v_q\}} \propto \frac{1}{dist(v_p, v_q)} \cdot exp\left(-\frac{(I_{v_p} - I_{v_q})^2}{2\sigma^2}\right), \tag{12}$$

which makes $B_{\{v_p,v_q\}}$ large for discontinuities between similar pixels/voxels (i.e., $|I_{v_p} - I_{v_q}| < \sigma$) and small for pixels with very different properties (i.e., $|I_{v_p} - I_{v_q}| > \sigma$).

Suppose a graph with numerous terminals illustrating every target object in an image/volume, and the hard constraints imposed by a user by properly placing seeds of distinct colors, the graph-based method can be expanded to simultaneous segmentation of multiple objects. The equivalent cost function is a straightforward generalization of Eq. (10) to the case of multi-label setting,

$$E(C) = \lambda \sum_{v_p \in \mathcal{V}} R_{v_p}(C_{v_p}) + \sum_{\{v_p,v_q\} \in \mathcal{N}} B_{\{v_p,v_q\}}(C_{v_p}, C_{v_q})\delta(C_{v_p}, C_{v_q}), \tag{13}$$

where $C \in \{\mathcal{O}_1, \mathcal{O}_2, \cdots, \mathcal{O}_N\}^M$ denotes the label assignment of a multi-object image, and $B_{\{v_p,v_q\}}(C_{v_p}, C_{v_q})$ is the discontinuity penalties between a pair of neighboring pixels/voxels according to their associated labels. A multi-channel graph cut with minimum cost that partitions the terminals into separate parts by splitting those edges with minimal cumulative cost corresponds to an optimal multi-object image segmentation.

It is worth noting that a directed graph can be easily generalized from an undirected one by allowing different discontinuity penalties (directed costs) for two cases: $p \in \mathcal{O}$, $q \in \mathcal{B}$ and $q \in \mathcal{O}$, $p \in \mathcal{B}$, i.e., two distinct node pairs: (v_p, v_q) and (v_q, v_p).

2.3. Stochastic formulation

Many problems such as image segmentation involve allocating a label to every pixel/voxel, which can be effectively interpreted as energy minimization in respect with Markov random fields

DOI: 10.1002/cnm

(MRFs) [14, 15], a basic type of stochastic process. The resulting energy function is made up of two parts: data term (also known as unary term) penalizing results which are not compatible with the observed images/volumes, and smoothness term (also called pairwise term) enforcing spatial coherence. In other words, unary term represents data likelihoods, while pairwise term implies data smoothness. For simplicity, let x_p be the label associated with pixel p in a 2D image, $\boldsymbol{x}$ denote all the pixel-label allocations, M represent the total number of pixels, the MRF energy function can be defined as,

$$E_{MRF}(\boldsymbol{x}) = \lambda E_s(\boldsymbol{x}) + E_d(\boldsymbol{x}), \tag{14}$$

$$= \lambda \sum_{\{p,q\}\in\mathcal{N}} V_{pq}(x_p, x_q) + \sum_{p\in\mathcal{V}} D_p(x_p), \tag{15}$$

where $\mathcal{V}$ denotes the pixel set, $\mathcal{N}$ represents all the neighboring pixel pairs under a traditional 4-connected neighborhood system[†], $\{p, q\}$ is an unordered group of neighboring pixel pairs, $E_d(\boldsymbol{x})$ and $E_s(\boldsymbol{x})$ stand for the data energy and smoothness energy respectively. In the MRF scheme, the (negative) log likelihood of the measurement noise mainly forms the data energy $E_d(\boldsymbol{x})$, while the negative log likelihood of the prior forms the smoothness energy $E_s(\boldsymbol{x})$. Usually there are no restrictions on $E_d(\boldsymbol{x})$, however, $E_s(\boldsymbol{x})$ is particularly adopted for specifying the assumption of smoothness on pixel labeling. A certain V_{pq} corresponds to some specific assumption. A more regulated expression of $E_s(\boldsymbol{x})$ can be given by,

$$E_s(\boldsymbol{x}) = \sum_{\{p,q\}\in\mathcal{N}} \omega_{pq} \cdot V(|x_p - x_q|), \tag{16}$$

where ω_{pq} denotes the spatially varying pair weight and $V(|x_p - x_q|)$ stands for a non-decreasing function of the dissimilarity between pairing labels. A common choice of such a function is defined as (the Potts model),

$$V(|x_p - x_q|) = 1 - \delta(|x_p - x_q|), \quad \text{here } \delta(\Delta x) = \begin{cases} 1, & (\Delta x = 0) \\ 0, & (\Delta x = 1) \end{cases}. \tag{17}$$

This function uniformly penalizes any pair of pixels with distinct label allocations.

The minimization of the energy functionals discussed above can be typically solved by gradient descent optimization algorithms [16], combinatorial approaches [12], primal-dual methods [17, 18], and graph cuts [19], which will be further described in the next section.

3. APPROACHES TO ENERGY MINIMIZATION

Given the formulated energy function, the minimization is a numerical procedure to find an optimal solution (local or global) on the potential energy surface so as to attain the preferred results (e.g., segmentation, registration), which could be accomplished in a number of means (see Table I).

[†]The extension to other grid topologies such as 8-connected in 2D and 26-connected in 3D is straightforward.

DOI: 10.1002/cnm

Table I. Examples of energy minimization techniques.

Method	Continuous	Discrete	Global	Local	References
Newton-Raphson	Yes		Possible	Yes	[20, 21, 22, 23, 24, 25]
Gradient descent	Yes		Possible	Yes	[26, 27, 28, 29]
Conjugate gradient	Yes			Yes	[30, 31, 32]
Proximal gradient	Yes		Yes	Yes	[33, 34, 35, 36, 37]
Coordinate descent	Yes			Yes	[38, 39, 40, 41, 42]
Genetic algorithm	Yes		Yes		[43, 44, 45, 46]
Graph cuts		Yes	Yes		[47, 48, 49, 50, 51, 52]
Belief propagation		Yes	Yes		[53, 54, 55, 56, 57]
Tree-reweighted message passing		Yes	Yes		[58, 59]
Linear programming		Yes	Yes		[60, 61, 62, 63, 64, 65, 66]
Maximum margin learning		Yes	Yes		[67, 68, 69, 70, 71]
Simulated annealing		Yes	Yes		[72, 73, 74]
Iterated conditional modes		Yes		Yes	[75]
Minimal surface	Yes	Yes	Yes	Yes	[76, 77, 78, 12, 79, 80, 81]

Note: Continuous, Discrete, Global, and Local: types of optimization.

3.1. Continuous optimization methods

This class of optimization techniques seeks to minimize an appropriate energy function on a continuous space, among which the earlier work was proposed in 1980's [1, 82]. Recently, level-set representation has been mainly applied for continuous segmentation [83, 84, 85], where a shape can be embedded in respect with a zero-level-set of a function, and the shape priors are described on the space of the embedding function. The optimization in regard to deformation of shapes means minimizing an energy function. Among many other optimization approaches, gradient methods (e.g., gradient descent [26, 27, 28, 29], conjugate gradient [30, 31, 32], proximal gradient [33, 34, 35, 36, 37]) are particularly popular in the field.

3.1.1. Newton-Raphson method The Newton-Raphson/Newton's method [23, 24] was developed according to expanding a prospective energy surface at the present geometry utilizing Taylor series, depending upon its first derivative (i.e., slope/gradient) along with second derivative (i.e., curvature). The evaluation of these derivatives at every iteration makes the Newton-Raphson method computationally expensive, particularly in case of a multi-dimensional prospective energy surface on which there exist numerous directions for calculating the curvatures as well as gradients. Fortunately, this method normally needs the fewest iterations to find the minimum, i.e., rapid convergence. In general, the convergence is quadratic. The Newton's method could be applied for finding the optimum of an energy function with a zero-value derivative, assuming that the function approximates to a quadratic in regions close to the optimum, and utilizing the derivatives of first- as well as second-order for finding the stationary point. In higher dimensions, the gradient together with Hessian matrix of the second-order derivatives of the function undergoing minimization are adopted in the Newton's method.

The quasi-Newton method is an algorithm derived from Newton's method for finding the local optimum or stationary point of a function with a zero-value gradient. The most prevalent approaches include the symmetric rank one (SR1) technique [86, 22], the Berndt-Hall-Hall-Hausman (BHHH) technique [20], and the ubiquitous Broyden-Fletcher-Goldfarb-Shanno (BFGS) technique [87, 25] along with its limited-memory expansion: L-BFGS technique [88, 21]. The SR1 algorithm updates

DOI: 10.1002/cnm

the second derivatives (Hessian matrix) with regard to the gradient estimated at two points, which keeps the matrix's symmetry without ensuring its positive definiteness being preserved. The BHHH approach can guarantee that the iterative procedure converges under particular circumstances. The BFGS framework is an iterative scheme to tackle nonlinear optimization tasks with no constraints, which attempts to attain a stationary point of a function that can be continuously differentiated for two times. Similarly, it does not assure of being converged except that the function can be quadratically expanded nearby the optimum utilizing Taylor series, but BFGS can perform fairly well even in case of non-smooth optimization. The L-BFGS strategy approximates BFGS by only involving low memory, which is extraordinarily suitable for optimization tasks with numerous variables. In contrast to BFGS that saves a compact approximation of the reversed Hessian matrix, L-BFGS uniquely keeps several vectors which implicitly depict such an approximation.

Compared with Newton's method, the Hessian matrix of second derivative in quasi-Newton method is not directly evaluated, but renewed via examining consecutive gradient vectors, and the reversed Hessian matrix for steering the line search is usually generated using an approximation instead of being estimated which is expensive in almost every respect. However, the memory required for operation in quasi-Newton method is generally more than Newton's method.

3.1.2. Gradient descent method Gradient descent (called steepest descent as well) is a first-order iterative optimization method, which is used for discovering the function's local minimum by shifting in the negative orientation of the present gradient of the function. This procedure converges ultimately at the location with a zero-value gradient. Without requiring the estimation of second derivative, this scheme is considerably quicker per step than Newton's method, but normally needs more steps to reach the minimum. Due to its applicability and simplicity, in medical image analysis, gradient descent is the most prominent optimization technique used in various deformable models [1, 8, 89], where explicit differentiation is performed in accordance with the Euler-Lagrange equations. When these equations meet the expectations, it is assured of being converged to an optimal result.

In [26, 27], Tsai et al. applied the gradient descent algorithm to minimize a region-based energy function for segmenting the medical images. However, gradient descent algorithm suffers from slow convergence. In the existence of image noises, implementing the gradient descent process may cause unreliability in the deformations as time goes by. Moreover, the search space's complexity will increase if the total number of dependent factors (e.g., orientation, scale, location, shape) rises, so in this case the number of local minima can grow and it is also necessary to calculate more and more derivatives. To overcome the difficulty, Sundaramoorthi et al. [28] reformulated the traditional geometric active contour (GAC) models, where the gradient is redefined employing the inner product of Sobolev-type. As a result, the advantageous regularity properties are successfully induced into the gradient flows, thus make the deformations smoother over time, preventing some unacceptable local minima to a large extent. In [29], Bar and Sapiro presented a generalized Newton-type scheme in terms of more generic norms than the l_2-type norm. These norms are introduced in calculating the descent as well as in the relevant trust-region stabilization. In addition to quicker quadratic convergence, the proposed method achieved higher accuracy in the existence of noises.

A problem with gradient descent is that for those functions with valleys, the method zigzags as the gradients are almost perpendicular to the orientation of the local minimum in such areas. This is

Prepared using cnmauth.cls
DOI: 10.1002/cnm

similar to the case of being in a disc-like tube and attempting to remain in its lower position. In the event that the local minimum is at the tube's bottom, this will require much more time to achieve the minimum since the algorithm keeps jumping between the tube's two sides (zigzagging). In addition, it is crucial to select an appropriate step size so as to avoid divergence because of larger step size or slower convergence due to smaller step size. There are two options, one is to pick a fixed step size which ensures the algorithm's convergence regardless of where gradient descent starts (i.e., constant selection), the other is to take a separate step size per iteration (i.e., adaptive selection) [90, 91].

When the energy functional is convex, gradient descent gains global convergence, because all the local minima are global ones as well. However, problem may arise in case that the optimum is not within the convex optima set. In such a case, the gradient descent procedure can be integrated with an iterative projection strategy for the minimum calculation of an energy function [92]. For convex problems, Nesterov provided a straightforward refinement of gradient descent to speed up its convergence [93]. If the projection can be computed in an efficient way, gradient descent may be generalized to cope with constraints by involving a projection onto the constraint set, a particular example of the popular forward-backward technique (a proximal splitting method) for monotone inclusion problems [37].

Gradient descent can also be applied for higher-order optimization tasks. In accordance with the observation that an energy function or its part can often be differentiated, Ishikawa [94] introduced a strategy that enables the fusion move procedure [95] quite faster for potentials of higher-order. The presented method can yield the proposal labeling via shifting the present labeling by the energy gradients, i.e., the proposal labeling in every iteration can be produced utilizing an elementary gradient descent. Dissimilar to the standard gradient descent, the proposed algorithm can still effectively take the gradient using even a partial energy. For example, by utilizing the higher-order priors solely, its gradient can be achieved, and the labeling can then be shifted in that orientation, which speeds up the fusion moves. Another benefit in comparison with standard gradient descent is that it is possible to make the descent step very big, accelerating the optimization process especially in the beginning stage. The reason for both improvements is that the fusion move assures a decrease of the energy.

3.1.3. Conjugate gradient method Similar to the gradient descent method, the starting part of the search in the conjugate gradient algorithms [31, 32] arises in the opposite (steepest descent) orientation of the biggest gradient. To prevent the oscillating backwards and forwards problem occurred in the gradient descent procedure when it walks towards the minimum, the conjugate gradient strategy takes into account the previous orientation in the subsequent search, enabling the rapid movement to the minimum. The conjugate gradient method can produce an accurate result within a limited iterations, thus it is a direct approach theoretically. However, it is unstable even because of slight perturbations. Fortunately, it can also be viewed as an iterative technique since the algorithm provides increasingly better approximations to the actual result, satisfying the specified criterion of tolerance within a fairly small number of iterations.

Mostly, preconditioning is essential for making the conjugate gradient approach converge rapidly for solving linear systems. An important variant of such preconditioners is the inexact preconditioned conjugate gradient method [96] accomplished by combining the inward and outward iterations, where the preconditioned equations are worked out employing an inward iteration to

DOI: 10.1002/cnm

a predetermined accuracy. In some practical implementations, the preconditioners applied are of high complexity, potentially leading to inconsistent preconditioning that changes across iterations, thus resulting in a remarkable increment of the time to make the algorithm converge. To improve the convergence in such cases, Golub and Ye [97] formulated an inexact preconditioner to solve a symmetrical and explicit definite system. Applying a local connection between consecutive residual norms, they obtained a result that converges linearly. Furthermore, employing a global relation, the presented method is able to converge more linearly if the inward iteration can be resolved with good precision, suggesting that the stopping criterion for the inward iteration can be heuristically selected.

To preserve the optimal convergence property, Axelsson [98] proposed a variant of the conjugate gradient method where the search direction vectors are explicitly orthogonalized. The algorithm is referred to as flexible conjugate gradient [99]. It is particularly developed to adapt to variable preconditioners. In practice, flexible conjugate gradient is required to be incorporated into a truncation/restart scheme. In [99], Notay developed a hybrid scheme, which integrates the benefit of pure truncation with the advantage of pure restart. He also theoretically demonstrated that the presented work can converge independently from the preconditioner variations provided that the variations are slight enough. Under specific conditions, its convergence rate is comparable to that of the truncated conjugate gradient algorithm or the conventional one, while permitting rather big variations with almost no penalty. In some other circumstances, flexible conjugate gradient achieves better performance than traditional conjugate gradient, whereas the truncation requirement restricts the variation levels entitled.

The nonlinear conjugate gradient method [100] is a generalized version of the conjugate gradient approach to tackle nonlinear optimization issues, which discovers the local minimum of a nonlinear function that can be differentiated for two times at the minimum. It performs a line search in the opposite orientation of the gradients until achieving the minimum.

Different from Newton-based methods (e.g., Newton-Raphson algorithm and quasi-Newton method) that compute the Hessian matrix of second derivative except for the gradient, conjugate gradient methods need much less estimation in every iteration.

3.1.4. Proximal gradient method Proximal gradient (forward backward splitting) methods [33, 37] are gradient-based algorithms suitable for convex optimization problems, where the objective function consists of two convex functions, one is differentiable with Lipschitz continuous gradients while the other one may be non-differentiable (non-smooth). They are called proximal because each non-smooth function is engaged by means of its proximity operator, which can be viewed as a generalized form of a projection [33, 91]. An essential method relevant to proximal techniques is the Moreau decomposition where the identity operator is decomposed into two proximity operators [33]. Similarly, we can consider it as the generalized version of a vector space's orthogonal decomposition. Typical proximal algorithms include iterative shrinkage/thresholding algorithm (ISTA) [101, 102, 103, 90], Landweber iteration [104, 105], alternating projection [106, 107], alternating direction method of multipliers (ADMM) [108, 109], and split Bregman method [110].

ISTA is one of the most popular techniques for tackling convex unconstrained optimization problems, in which every iteration includes matrix-vector multiplication followed by a shrinkage/soft-threshold step [101, 102, 103, 90]. This iterative shrinkage-thresholding scheme can be treated as a natural expansion of the conventional gradient method, which is interesting because

DOI: 10.1002/cnm

it is simple but suitable for tackling problems of large-scale with high-density matrix data. As a first-order method where the optimization depends upon function value as well as gradient evaluation, ISTA is a preferred option in practice for large-scale problems in many instances. However, the linear observation operator dominates the convergence rate of ISTA, which may become fairly slow in case of ill-conditioned/ill-posed operator [102, 33]. A variety of approaches have been proposed to accelerate the performance of ISTA. The authors in [34] developed an interesting two-step ISTA algorithm called TwIST, which converges much quicker than ISTA for problems that are ill-conditioned. Regarding of numerous regularizers that are convex but non-quadratic, it was proven that TwIST is able to converge to its objective function's minima, for an appropriately chosen band of values of those parameters formulating the method. Its higher effectiveness than ISTA has been experimentally exhibited on many linear inverse problems like image restoration and deconvolution. In [35], Elad et al. presented a method accelerating ISTA for identical problems as shown in [34], which applies the sequential subspace optimization technique and computes the following iteration through minimizing a function on an affine subspace that is spanned by the present gradient and at least two earlier iterations. However, for both methods [34, 35], global non-asymptotic convergence rate has not been achieved. To overcome the difficulty, Beck and Teboulle [36] proposed a fast ISTA (FISTA) that maintains ISTA's simplicity in computation and significantly increases its convergence rate. They studied the global non-asymptotic convergence rate and the efficiency of ISTA algorithm assessed by function values. This common analysis is able to cope with an objective function with arbitrary continuous convex non-smooth regularizer along with convex smooth function. The initial numerical solutions of the image deblurring tasks using wavelets show that FISTA performs rather better than either ISTA or TwIST. It is worth noting that all of these methods [34, 35, 36] depend upon generating the subsequent iteration in regard to at least two earlier iterations and the present one.

The Landweber iteration is an iterative method for solving linear inverse problems that are ill-posed [104], and has been generalized for tackling nonlinear ill-posed problems involving constraints [105]. It may be treated as a particular version of many other generic techniques (e.g., gradient descent) [37]. Many variants of the Landweber algorithm have been developed in the literature. For example, the high-dimensional least squares problem that is sparsely constrained can be tackled employing the Landweber method [111] (a special case of projected gradient descent). To deal with the optimization of a weighted least squares problem that is convexly constrained, Johansson et al. [112] exploited the projected Landweber algorithm that copes with convex constraints via enabling oblique projections instead of conventional orthogonal projections onto closed convex sets. They also proved the convergence of the oblique-projected Landweber method mathematically, which can be accelerated by choosing an appropriate preconditioner, in comparison with standard projected gradient search.

The alternating projection (also called projections onto convex sets) method [106, 107] intends to detect a crossing point from two closed convex sets, where a point is interchangeably projected onto the two sets. Specially, for circumstances that the sets are affine spaces, the algorithm converges to the crossing point and also to the orthogonal projection of the original point. The alternating projection scheme converges at a linear rate [113]. The Dykstra's projection algorithm [114, 115] is a variant of the alternating projection method. If the number of crossing point is larger than one, the classical alternating projection randomly yields a point in the intersection part of the two sets, while

DOI: 10.1002/cnm

Dykstra produces some specified point, that is, the projected initial point onto the intersection part. Extensions of the alternating projection method examine circumstances with several sets, nonconvex sets, or those producing quicker convergence. In [116], Lewis and Malick developed a nonconvex alternating projection approach and proved its local linear convergence. More variants, applications, along with extensions of alternating projection can be found in [117, 118]. For the cases involving circulant system, the Landweber iteration adopting a specific step size is identical to a variant of the alternating projection technique [119]. Additionally, the projection methods converge faster than the Landweber iteration, as they make it easier to integrate additional constraints into the standard iteration.

ADMM [108, 109] is a variation derived from the augmented Lagrangian framework, which applies limited up-to-date information to the paired variables. Instead of iteration till achieving convergence, the algorithm continues to update the dual variable straightforwardly and repeat this procedure subsequently. The ADMM technique may be treated as an implementation of the Douglas-Rachford splitting method, which is an example of the proximal point scheme [109]. ADMM can be utilized for tackling separable convex programming problems [120]. The authors in [121] provided a survey of ADMM including variants and extensions. They argued that the algorithm is quite suitable for a variety of distributed convex optimizations, particularly for large-scale cases emerging in machine learning and statistics. In order to lower the computational cost of inner-loop coordination, Tosserams et al. applied the ADMM for updating penalty criteria following one inner-loop iteration [122]. Experimental results demonstrated that the total amount of sub-problem minimizations and the costs could be largely decreased by 10 to 1000 times. The convergence rate of ADMM has been discussed by many authors in the literature. For example, Boyd et al. [121] proved the objective and residual convergence of ADMM. Hong and Luo [123] analyzed the ADMM's global linear convergence in optimizing the summation of at least two non-smooth functions that are convex and separable under linear constraints.

The split Bregman scheme [110] is an approach based on Bregman iteration [124] to tackle a wide range of general l_1 regularized optimization problems. Specially, it can effectively solve those problems with total-variation (TV) regularizations. The method can also be viewed as a reinterpretation of the ADMM exceptionally adapted to l_1 problems. An advantage of this method is that the continuation, enforcement, and regularization of inequality constraints is not needed. Moreover, the algorithm can be easily parallelized because it makes extensive implementation of the Fourier transform and Gauss-Seidel techniques. In comparison with those second-order approaches which need exact descriptions of Hessian matrix, its requirement of memory is relatively low. Thus, the split Bregman framework can be used to provide very efficient solutions for many problems such as TV denoising, compressed sensing for image reconstruction from Fourier coefficients, and convex image segmentation [125, 126]. In [127], the author proposed two improved linearized Bregman methods to tackle the basis pursuit (BP) and relevant sparse minimization issues. The linearized Bregman technique was interpreted as a gradient descent approach on the dual task, thus acceleration schemes for optimization strategies relying upon gradients like line search may be used for tackling the dual task. The experimental results demonstrate that the improved methods are mostly quicker than linearized Bregman involving 'kicking'. Similarly, Yang et al. utilized the split Bregman strategy to the dual formulation of the speedy l_1 minimization problem emerging in compressed sensing [128]. Compared to Yin's gradient descent-based algorithms [127], the proposed dual split

DOI: 10.1002/cnm

Bregman method is more efficient, especially for cases containing under-determined matrix. Further accelerating strategies utilizing the equivalence between linearized Bregman and gradient descent have been reported in [129]. In addition, the convergence rate analysis of the split Bregman approach may be obtained in [110, 130, 128]. Esser [131] explored the connection between split Bregman and ADMM. Therefore, the inexact minimization and alternating step applied in split Bregman for instances where the methods are comparable can be validated using the existing convergence theory for ADMM.

Proximal methods provide effective optimization procedures for tackling the variational tasks that define the l_1 regularization schemes [102, 36]. These approaches are able to manipulate a variety of machine learning techniques, termed as structured sparsity regularization [132]. With an appropriate perturbation of the initial objective function (e.g., adding a tiny purely convex term), it is possible to improve the computational performance with no effect on the produced result with regard to its prediction power [132].

3.1.5. Coordinate descent method Coordinate descent [38, 39, 42] is a non-derivative multi-dimensional optimization scheme, in accordance with the strategy that an nD optimization task can be separated into n independent 1D sub-tasks solved by a suitable 1D optimization algorithm (e.g., dichotomy method and golden section method [133]). The local minimum of a function can be obtained by performing line search alongside a coordinate orientation at the present point in every iteration. Rather than changing the descent orientation in respect with gradients, the descent orientation is fixed in the very beginning. During this procedure, nonidentical coordinate orientations are cyclically iterated, one at a time while all other orientations remain unchanged, minimizing the objective function in regard to that coordinate direction. Coordinate descent may be considered as a particular example of block coordinate descent [134], which splits the coordinates into a few blocks and iteratively minimizes the function on the basis of one coordinate block while the other coordinates keep no change. In [40], the authors estimated the multi-shape model's parameters employing the coordinate descent framework for minimizing an energy functional to segment medical images utilizing the mutual information.

Generally, it is tough to tackle the linear least squares problem involving separable and non-quadratic regularization by exploring traditional iterative techniques. In [41], Elad devised a straightforward iterative strategy called parallel coordinate descent, which can be employed as a successful approximate solution for the basis pursuit denoising (BPDN) problem. Its performance can be further sped up by utilizing a serial subspace optimization method without increasing much complexity [35].

Coordinate descent algorithm generally works well for separable functions. For a non-separable function, it may have difficulty in discovering the optimum within limited iterations. To speed up the convergence rate, Loshchilov et al. [133] developed the adaptive coordinate descent scheme, an expansion of the coordinate descent approach to non-separable optimization problems. This approach applies adaptive encoding [135] that is based on the powerful covariance matrix adaptation (CMA) [136, 137] to progressively learn an appropriate transformation of the coordinate system, so that the up-to-date searching coordinates generated utilizing principal component analysis (PCA) are largely decorrelated in respect with the objective functions. The adaptive coordinate descent algorithm is invariant to monotonous function transformation (scale-invariance) and orthogonal

DOI: 10.1002/cnm

search-space transformation (rotation-invariance). In addition, the utilization of a suitable coordinate system makes it perform better than coordinate descent on a non-separable function. Its convergence is much quicker than coordinate descent and is as good as a method depending upon gradients. The time-complexity of this novel technique presents linear property and it fairly suits nonlinear optimization problems of large scales as well, by extending the adaptive encoding scheme to the nonlinear circumstances applying kernel PCA. Moreover, comparison-based surrogate models [138] can be exploited for maintaining the invariance property of the adaptive coordinate descent method.

3.1.6. Genetic algorithm-based method Genetic algorithms (GAs) have been implemented for minimizing the energy functionals as well, although the global optima are not formally assured [44, 45, 139]. With GAs, the traditional active contour model was extended for direct minimization of the basic energy functional [140]. Tohka [44] introduced a hybrid algorithm combining GA with greedy algorithm [141] for the global optimization of deformable surface model represented by simplex meshes. In [45], the authors developed a GA-based technique for explicit active contours. In this work, Fourier descriptors are adopted and parallel GAs are employed for accelerating the optimization. Hill and Taylor [43] incorporated GA with active shape models, in which the maximized objective function indicates the similar nature between the gray-scale intensities of the searching objects and the training objects. Instead of using the distinctive gradient descent optimizer, the authors in [46] explored GAs to optimize deformable models that investigate the medial-based, localized, and nonconvex shape variations in an explicit fashion. By evolving various models simultaneously, the application of GA alleviates the weaknesses of deformable models relevant with local minima, pose estimation, as well as model initializations. In addition, the search-space's size is decreased through restricting the evolution.

3.2. Discrete optimization methods

Algorithms for discrete energy minimization play an important role in medical image analysis, which mainly includes graph cuts [47, 48], belief propagation (BP) [54, 56, 15], tree-reweighted message passing (TRW) [58, 59], linear programming (LP) [61, 142, 143], and maximum margin learning [67, 71].

3.2.1. Graph cuts method Many optimization techniques like gradient descent are prone to local minima that could be arbitrarily distant from the genuine optima, therefore provide no promise of the optimality of the produced results. Recently, a number of approaches [12, 79, 13, 49, 144, 9, 92, 145] attempt to achieve the global minima of an energy function in case that the local minima are not satisfactory. These techniques can be grouped into three categories: min-path, mini-cut, together with convex approximation. Min-path methods [146, 147, 148, 149] are constructed in respect with Dijkstra's algorithm [150] to search for a path with smallest distance in an undirected graph whose edges are associated with non-negative weights, while convex approximation [144, 9, 92] relaxes the potential shape models from nonconvex spaces to convex spaces, thus defines convex energy function that could be globally optimized. In convex relaxation, the energy function, the shape space where they are minimized, and the relaxed shape space should be all convex.

Graph cuts (mini-cut, computed by max flow algorithms) [151, 152, 47, 12, 48, 49, 51] are a global minimization technique for a prevailing energy function, a specific instance of calculating a

Prepared using cnmauth.cls
DOI: 10.1002/cnm

geodesic over a Riemannian space with an image-based metric. In [47], Boykov et al. introduced two popular graph-cut techniques: α-expansion moves and α-β-swap moves. Practically, the great efficiency of the state-of-the-art expansion algorithm has been largely shown, however, it generally applies to MRFs with metric potential functions. On the other hand, the swap algorithm can handle more generic energy functions, as it is applicable to nonmetric potentials as well. Its drawback is that it seems to be less effective than the expansion method. The two popular graph-cut methods both rely upon moves which only allocate every pixel two labels for selection. Their performance is particularly good for energies in which the potential MRFs have Potts prior, corresponding to the assumption of a genuine piecewise-constant labeling. However, their performance may not be good for more generic priors. To overcome the difficulty, Veksler [52] developed a new graph cuts-based algorithm called range moves for those MRFs involving truncated convex priors, which correspond to the assumption of piecewise smoothness. To achieve a high-quality approximation, the range-move algorithm allows every pixel to choose from at least two labels. In order to yield a more prevailing perspective of the graph cuts-based combinatorial optimization techniques, Komodakis and Tziritas [153, 50] presented a novel theoretical scheme for the approximate optimization of MRFs with metric energy functions as well as non-metric ones, utilizing the duality principles of linear programming. The novel primal-dual algorithms could deliver instance-based sub-optimal bounds in any case which includes discrete MRFs with any potential functions. In practice, such bounds have been shown to be quite solid (near 1), indicating that the produced results are optimal to a certain extent.

Graph cuts may only be utilized to submodular energy functions [48]. The produced optimal results are discrete approximation of continuous formulations [12]. These algorithms are based on combinatorial optimization. They are well suited for image segmentation, where the pixels or voxels can be represented by the graph nodes and the neighborhood relationship between pixels or voxels can be represented by the edges of the graph. A minimum-cost cut partitioning the nodes of the graph generates a segmentation result of a potential image/volume, which is optimal on the basis of edge weights.

3.2.2. Belief propagation method The belief propagation (BP) or sum-product message passing approach is a technique utilizing message passing to carry out the inferring process on graphical models (e.g., MRF, Bayesian networks), which evaluates the marginal probability distribution of every unobserved vertex in a graph, subject to observed vertices. It was first formulated on trees, then poly-trees as an extension. The BP scheme is a valuable approximation method for generic graphs [53]. In [154, 155], the authors formulated the stereo matching task as a Markov network that is made up of three coupled MRFs, and applied the Bayesian BP scheme to achieve the maximum a posteriori (MAP) evaluation in the Markov network. Sarkis and Diepold [156] proposed a method utilizing the BP algorithm to roughly calculate the disparity map of sparse images. The proposed sparse BP possesses a higher efficiency than the classic BP in terms of memory consumption, as only the nonuniform samples together with their 4-connected neighborhood pixels are stored and subsequently handled from the left images. The authors in [55] extended earlier theoretical solutions on tree reweighted BP (TRBP) [157, 58, 59] to secure a global minimum for the stereo energy function, accounting for potential ties in the beliefs. Like linear programming relaxation [158], TRBP is assured of obtaining a global minimum in case of non-fractional beliefs.

DOI: 10.1002/cnm

Loopy BP (LBP) [54, 56] is a variant of BP algorithm applied to a graph with cycle (loop). Generally, it is not assured that the LBP algorithm can reach convergence. To improve its efficiency, Felzenszwalb and Huttenlocher [159, 57] presented numerous ways including the distance transform method. In [14, 15], Szeliski et al. performed two distinct LBP variants: BP-M (also known as max-product LBP) and BP-S (also known as LBP application derived from TRW-S) to compare their performance in energy minimization. As a result, BP-M performs fairly good.

3.2.3. Tree-reweighted message passing method Tree-reweighted message passing (TRW) was first presented by the authors in [58] as a novel message passing mechanism for energy minimization. Compared with BP, the TRW technique could be utilized to much more energy functions. However, it does not always converge, and is not assured of increasing the energy's lower bound with time. To solve the problem, Kolmogorov [59] developed an improved TRW, that is, sequential TRW (TRW-S), where the lower bound value is promised not to become smaller, resulting in some convergence. In comparison with original TRW, the TRW-S refreshes the messages sequentially instead of in parallel order. Compared to conventional message passing schemes like BP, another advantage of TRW-S is that its implementation requires only half of the memory, but it achieves a much better performance than the original TRW [58] and the standard max-product BP [15]. Moreover, in contrast to graph cuts [47], TRW-S can achieve a slightly lower energy in stereo cases involving Potts interactions.

3.2.4. Linear programming method Linear programming (LP) or linear optimization [61, 142, 143] means optimizing a linear objective function over a feasible region, in regard to a collection of linear equality as well as linear inequality constraints. LP problems can be tackled by basis-exchange algorithms (e.g., simplex [60, 160], criss-cross [161, 62], and conic sampling [66]), and interior-point methods (e.g., ellipsoid [162], projective [163], and path-following [164]).

To solve the LP problems utilizing the popular simplex method [60, 160], the standard form describing an LP problem must be converted into augmented version that brings in slack variables with non-negative values for the substitution of inequality constraints with equality constraints. The simplex algorithm then solve the LP issues through yielding a viable result at a poly-tope vertex and moving down the poly-tope edges to its vertices associated with declining function values till obtaining a minimum/maximum. The shape of the poly-tope is described using the constraints imposed on the objective function. In practice, 'stalling' often occurs where numerous pivots are generated without augmenting the function values [143], while 'cycling' sometimes happens where the deterministic pivoting rules yield a loop without limit [165]. A variety of state-of-the-art pivoting rules [166, 61, 62, 165, 143] have been designed for preventing cycling and thus ensure the termination of the simplex procedure. The popular simplex method is normally rather efficient and can assure of achieving a global solution. However, it may have worst-case exponential time-complexity [167] or polynomial average-case time-complexity [168] on certain problems. To further improve the efficiency, researchers developed the revised simplex algorithm [169, 170, 143]. In addition, advanced simplex techniques such as parametric self-dual simplex [171] solve the LP problems through converting them utilizing an approach alike Big M [171]. Despite its exponential time-complexity in theory, the tremendous efficiency of the simplex approaches in practical applications indicates the possibility of simplex variants running in polynomial time or

DOI: 10.1002/cnm

more efficiently. On the other hand, the simplex method along with its variants solve the LP tasks in a way that walks alongside the poly-tope edges, vertex by vertex, implying that the maximal edge number between poly-tope vertex pairs restricts the simplex methods' performance in theory.

The criss-cross method [161, 62] is another basis-exchange pivoting technique, but never cycles on linear programs. Compared with the two-phase simplex algorithm, it is simpler with only one phase, and has no requirement of maintaining the feasibility. Moreover, the criss-cross approach is a combinatorial algorithm for linear programming, which selects two parameters: one for entering, another for leaving, only taking into account the coefficients' signs instead of the real-number ordering of them. Like Dantzig's simplex method [60, 160], its time-complexity is not polynomial.

Similar to the simplex and criss-cross methods, the conic sampling algorithm [66] moves between vertices as well. However, it interchanges several bases each time with no restriction of moving alongside the poly-tope edges. When arrive at a vertex, the conic sampling method arbitrarily selects from a group of vertices associated with increased objective values, and moves forward alongside the predetermined direction until encountering a newly imposed constraint. Such a process is repeated till there is no more direction for improving the objective function and satisfying the bounding constraints, implying optimality. Such a stochastic optimization procedure is straightforward but effective, especially for highly constrained, sparse LP problems. Furthermore, it could be reshaped and efficiently utilized to circumstances such as some quadratic programs that evaluate projections onto poly-topes. However, the conic sampling's worst-case time-complexity is still exponential.

In comparison with the simplex methods that travel across the edges between neighboring vertex pairs of the poly-tope for discovering an optimal result, the interior-point methods [172, 173, 171] walk alongside the strict inner part of the viable region and seek a result in an iterative manner. If the present result meets a predetermined accuracy, they terminate and possibly utilize alternative approaches, moving downwards the closest vertex and attaining a precise result. The ellipsoid algorithm [162] is a scheme first introduced by Khachiyan for tackling the linear programs to a expected precision in polynomial time, which achieves an ellipsoid of minimum volume containing the overlapping area of the viable region and the half-space that require the objective value with no increment. Khachiyan's algorithm proves the solvability of LP problems, however, its efficiency is not as competitive as the simplex method. Inspired by Khachiyan's algorithm, Karmarkar proposed a projective method [163] in which a logarithmic barrier function is adopted for replacing the constraints and the feasibility along with the optimality are maximized at the same time. The projective algorithm is assured of tackling the LP issues with polynomial time-complexity. Unlike the ellipsoid algorithm, it is practically quicker than the simplex algorithm, and motivates significant attention of interior-point techniques. Other variants of the interior-point methods include the simplified primal affine-scaling method [174] and the popular path-following method [164].

For the MRF optimization, such a class of approximation algorithms [63, 64, 65] first systematically express the problem as a natural integer program and tackle the integer program utilizing an LP relaxation, and subsequently employ an arbitrary rounding scheme for yielding a near-the-optimum-integer result. To ensure a suboptimal solution, these techniques mostly suppose a metric MRF potential function.

Linear-fractional programming (LFP) [175] is an extension of the LP, and accordingly its objective function is a proportion of two linear functions. When the denominator is a constant

DOI: 10.1002/cnm

function with the value of one everywhere, LFP becomes LP. The LFP problems can be tackled applying a variant of the simplex method [61, 176] or the extended criss-cross strategy [177]. In fact, linear programming has motivated numerous essential concepts in the optimization theory like decomposition, convexity as well as duality.

3.2.5. Maximum margin learning method Maximum margin Markov networks [67] was proposed for tackling the supervised classification problems with multiple labels and multiple classes, which integrates the benefits of the support vector machines (SVM) and the graphical model. Image segmentation problem can be modeled utilizing random fields like Markov random fields (MRFs) and conditional random fields (CRFs). The joint probability described by them is generally articulated with an exponential energy, which usually contains two parts: node energies and interaction energies, modeling every random variable and dependencies between adjoining variables, respectively. Traditionally, the energies are linearly depicted on the basis of unary parameter and pairwise parameter. Maximum margin learning has been recently implemented for shaping such a model discriminatively using the ground truth data for segmenting new images. The idea is to evaluate the parameters of an energy functional, which enables the labeled samples to achieve the smallest energy by the biggest potential margin.

Given a labeled training set, Anguelov et al. [68] applied the maximum margin framework to optimize the parameters of a MRF model by finding the optimum compromise between edge and node features. Such a training strategy produces the global optimum weights. Following the work on maximum margin estimation for probabilistic models, the authors in [69] proposed an extended framework for structured prediction models, based on the max-margin criterion that produces a substitute of those probabilistic evaluation algorithms relying upon likelihood, via focusing on the robustness of a model's decision boundary. The learning algorithms defined in this framework can efficiently detect the scoring function parameters, therefore the highest scoring structures approximate to the expected structures on the training data. In [71], Szummer et al. reported a method of high efficiency to learn multiple optimal parameters of random field models, with reference to the maximum margin network and the structured SVM. They applied graph cuts to precisely maximize the margin learning of parameters for sub-modular CRF and MRF, and ensured the generalization to newly collected images through a big margin regularizer. In addition, the authors addressed acquiring parameters of the non-submodular cases utilizing α-expansion that generates nearly optimal results. These works suppose that an individual group of optimal weights is applicable for all the testing images. However, the weights compatible with different images could be completely non-identical. This provides motivation for tackling the parametric problems throughout the time of testing. Kolmogorov et al. [70] learned the optimal parameters from the training data utilizing the maximum flow approach, based on which a parameter range is provided. For any image of the testing set, an individual optimal parameter is decided by a user from the given range. The efficiency of the described scheme is not high, but it can be potentially boosted by handling the parametric problems by means of techniques like min-path.

3.2.6. Other methods There are many other discrete approaches to energy minimization, including the well-known earlier algorithms like simulated annealing (SA) [72, 74] and iterated conditional modes (ICMs) [75]. SA is a stochastic optimization procedure which seeks an energy function's

DOI: 10.1002/cnm

global optimum in a large search space (mostly discrete space). Such an optimization process first melts the system to be optimized at a high temperature, and after that lowers the temperature with the help of slow stages till the system becomes frozen with no more variation. For every temperature value, the simulation ought to continue before the system arrives at a stable condition. On the other hand, ICM applies a greedy procedure to detect the local minima. Starting from an initial label estimation, for every pixel, ICM votes for the label that makes the energy function value drop to a large extent. By repeating such a procedure, rapid convergence can be ensured. However, the optimization results are susceptible to the labeling initialization. Except for the entirely synchronous option, ICM is comparable to instant freezing in SA.

3.3. Minimal surface method

Minimal surface is a surface locally minimizing the entire surface area under certain constraints, which is comparable to possessing a zero mean curvature. This description can be expanded to contain surfaces with constant mean curvatures. This is a convex optimization problem and there are a number of numerical methods to tackle the minimal surface problems. Chopp [76] introduced a method utilizing the level-set curvature flow model, where the mean curvature motions supply flows which locally evaluate the minimal surfaces. The primary strategy is to estimate a minimal surface through linking a surface with an identified boundary and making it proceed in respect with the mean curvatures. Caselles et al. [77, 78] extended the geodesic active contours [8] (a popular geometric optimization technique for 2D image segmentation) to 3D segmentation, where the preferred boundaries are achieved by the computation of minimal surfaces within a space obtained from the prescribed images. It indicates that the object extraction is comparable to seeking a minimal surface via a deformation equation generated by its Euler-Lagrange. In [79], Appleton and Talbot developed an approach to globally calculate the minimal weighted surface for segmenting images via continual maximal flow in random Riemannian space involving scalar measures. The increased reliability and accuracy rate corresponding to globally minimal surface will benefit numerous current implementations employing graph cut or geodesic active surface (GAS). In [80], the discrete minimal surface task was described and tackled on a 3D weighted graph. A natural implementation of this minimal surface solution is the expansion of the 2D intelligent scissors or live wire scheme [178] to 3D segmentation. Recently, Grady [81] extended the 2D shortest-path segmentation techniques on weighted-graphs appropriately to 3D image segmentation by tackling the minimal surface problems directly through the calculation of a circulation network flow with minimal cost with regard to discrete differential geometry. To potentially alleviate the high sensitivity to initializations associated with active contour/surface methods [8, 77, 78], Boykov and Kolmogorov proposed a *geocuts* approach [12] to segment images which integrates geodesic snake models [8, 179] with graph cuts. The graph cuts approach is utilized to obtain globally minimal geodesic contours (2D) or surfaces (3D) in regard to randomized Riemannian metric in the event of multiple boundary conditions.

In [180], Bates et al. presented a surface-free energy minimization method, where the minimal molecular surface (MMS) is used for the modeling of complex biomolecules (e.g., proteins, DNAs and RNAs). Under biomolecular constraints, the MMSs are generated by minimizing the mean curvature of hypersurface functions in an iterative way according to the differential geometry theory. The proposed MMS is capable of handling internal as well as open cavities, and is typically free

DOI: 10.1002/cnm

of geometric singularities. As described in [181], the molecular surfaces can also be created by applying high-order fractional partial differential equations (PDEs). Employing the fast fractional Fourier transform algorithm, the proposed PDE transform provides a robust and efficient approach for molecular surface analysis.

3.4. Primal-dual method

An optimization problem can be considered from either the primal perspective or the dual perspective, which is called the duality. The way of solving a dual problem yields a (lower) bound to that of a primal one (minimization problem) [182]. Generally speaking, a duality gap exists between the solutions corresponding to these two problems. In case of convex optimizations, this gap equals zero for a constraint qualification circumstance, i.e., the optimal result to the primal task is provided by the dual one. There are many primal-dual algorithms proposed during the past decades for both continuous and discrete optimization problems. Here we cover a few of them.

The dual task generally means a Lagrangian dual [183, 182], while the others include the Wolfe dual [184, 185], the Mond-Weir dual [186, 187], and the Fenchel dual [188, 189]. In Lagrangian dual problems [183, 182], the Lagrangian is formed by employing the non-negative Lagrange multipliers for adding restrictions to the objective function, and after that finding the values of primal variables which make the Lagrangian minimized. The purpose of the dual task is to make the objective function maximized in accordance with some dual variables under restraints imposed on them. Under specific circumstances, the Lagrangian duality gap relies upon the primal problem's nonconvexity [190]. In [191], Giannessi discussed the advancements of the Lagrangian duality theory. The Wolfe dual problem [184, 185] is particularly a nonconvex optimization issue, where the objective function as well as constraints are both differentiable, and the equality constraints are generally nonlinear. It always reserves weak duality (the duality gap is larger than zero). The Mond-Weir dual [186, 187] is an approach for the nonlinear programming problem, which modifies the dual task for weakening the requirements of convexity. For convex and many generalized convex problems, both Mond-Weir dual and Wolfe dual are two special examples of a certain generalized Lagrangian dual [192]. The Fenchel's dual theorem [188, 189] provides duality results for constrained convex optimization problems, where the functions involved are a proper convex function with its convex conjugate and a proper concave function with its concave conjugate. When these functions together with the bounded linear mapping satisfy a constraint qualification, the Fenchel dual reserves strong duality, i.e., zero duality gap.

In the linear programming optimizations where the objective function as well as its constraints are both linear, the primal objective function linearly combines p variables. Each of the q constraints gives an (upper) bound on it. The primal solution is a p-component vector which maximize the objective function value depending upon the constraints. On the other hand, the dual objective function linearly integrates q dual variables whose values correspond to the upper limit of the q restrictions associated with the primal task, and accordingly each of the p dual restrictions imposes a lower bound on it. The dual vector multiplies specified values which control the primal constraints' positions. Altering the dual vector has the same effect as changing the primal (upper) bounds. For example, the shortest path issue in a positively-weighted graph may be expressed as a particular example of the min-cost flow, while the Dijkstra's method [150] is a primal-dual technique which tackles the dual task starting from zero. The well-known A* scheme is a primal-dual approach as

DOI: 10.1002/cnm

well [193], which processes the dual problem starting from -h (the constant heuristic h has a positive value).

In the nonlinear programming optimizations where the constraints may not be definitely linear, to ensure an optimal result, the primal objective function is generally required to be a convex one and has a dense sublevel set, i.e., Karush-Kuhn-Tucker (KKT) condition [194]. By enabling inequality constraints, KKT method for nonlinear programming extends the Lagrange multipliers strategy that permits equality constraints solely. In a standard nonlinear programming, the Lagrange dual function can be expressed by introducing Lagrange multiplier vector or dual variable. As a pointwise infimum corresponding to a collection of affine functions, the dual function has the virtue of being concave and produces a lower bound on the primal optimal value. The strong duality holds in case of convex primal problems with constraint qualifications (e.g., Slater conditions [195]). In [185], Geoffrion generalized the duality theorem of nonlinear programming to advance more theoretical and computational implementations. Unlike many other techniques, this approach does not employ the conjugate function principles, conventional mini-max theory, or differential calculus. It depends upon the comparatively fundamental theory of convexity and completely exploits the influential idea of some specified perturbation function. The author stated three elementary theorems (namely weak duality, strong duality, and optimality), and additionally established another six theorems relevant to the primal and dual problems. Later, Mangasarian [196] investigated the second- as well as higher-order duality issues in the nonlinear programming problems. It was shown that the weak duality is reserved between primal problems and their corresponding dual Mond-Weir problems, when the necessary and sufficient conditions are both satisfied, i.e., the primal is Mond-Weir weak duality invex [197]. Recently, Ivanov [198] introduced first- as well as second-order duality in the WD-invex problems, which are nonlinear programming problems under inequality constraints holding weak duality between primal and dual tasks. They proved the converse, strict converse, weak, and strong duality along with another several theorems for both first- and second-order WD-invex problems. Compared with Mangasarian's method, this second-order duality generalizes first-order duality, i.e., all the first-order duals are second-order duals for the primal problems. It was derived that every first-order duality result is complied with in second-order circumstances.

In the convex optimizations, the primal problem can be coupled with a dual one engaging its functions' conjugates. In case of primal objective functions being strictly convex, the dual problems can be specified with reference to a Moreau envelope, i.e., a smooth convex function that can be differentiated continuously with Lipschitz continuous gradients. In [91], Bauschke and Combettes mathematically explored the duality issue in convex optimizations. They addressed the interaction between primal and dual tasks in the circumstances of Fenchel duality and Lagrangian duality, and also the minimization with regard to both inequality constraints and equality constraints.

The idea of duality for handling the TV-based image restoration (ROF denoising) [199] was first introduced in [10], where Chan et al. employed Newton's method for tackling the primal-dual ROF problem. This primal-dual Newton's method has an advantage of quick locally quadratic convergence. Later, Chambolle [200] reported a semi-implicit gradient descent method with regard to the Lagrangian multipliers to tackle the dual problem of the ROF model. Due to its speedy global convergence and simplicity, this approach became very prevalent. A unified framework reformulating the primal-dual ROF problem to second-order cone programming was presented in [201], where the problem is sorted out by the primal-dual interior-point scheme, namely

DOI: 10.1002/cnm

MOSEK package. Essentially, the update direction in every iteration is solved utilizing Newton's linearization. Thus, the convergence rate of this approach is comparable to [10]. Recently, Zhu and Chan [17, 18] developed a hybrid primal-dual gradient strategy of high efficiency, which can be implemented for solving the TV and l_1 minimization problems. It explores the primal-dual information and swaps between the primal-dual formulations. Additionally, this decent-type framework has a fairly higher convergence rate compared to other popular approaches like [10, 200].

3.5. Multi-objective optimization method

In multi-objective optimizations, all the terms of the energy functional are optimized concurrently, where an optimal solution is achieved when no term can be refined without sacrifice of another one. In [202], such a solution is called Pareto optimal solution, and a group of them at the end of search is termed as tradeoff surface (or Pareto front). Among these solutions, no preference is desired except a sorting is supplied amid the objectives. Collette and Siarry [202] divided multi-objective optimization techniques into three groups: a priori optimization, progressive optimization, along with a posteriori optimization.

Hanning et al. [203] proposed a framework for image segmentation employing a piecewise approximation of images with continuous functions. This technique exploits the multi-objective optimizations to balance the approximation error and the total segment number. To further enhance the accuracy of image segmentation, Nakib et al. suggested an image thresholding algorithm recently [204] for concurrent optimization of a few segmentation criteria utilizing multi-objective optimizations.

3.6. Comparative studies

It is generally accepted that ICM usually has a low-quality performance [15] and the graph cuts method is a massive advancement over earlier techniques like SA [47]. To evaluate the performance of various optimization techniques, Tappen and Freeman [205] gave a comparison of graph cuts with LBP, while Boykov and Kolmogorov investigated the effectiveness of several max-flow schemes for graph cuts [206]. Szeliski et al. [14, 15] conducted a comparative study on four different energy minimization approaches including ICM, LBP, TRW, and graph cuts with regard to efficiency and accuracy. For the Potts model, graph cuts achieve a fairly good performance among which the expansion method performs marginally better than the swap one. TRW-S has a similar performance to graph cuts, however, it needs much more time to arrive at convergence. Regarding of highly connected graphs with considerably higher connectivity, Kolmogorov and Rother [207] compared LBP, TRW, and graph cuts, and observed that graph cuts perform rather better than the others. In addition, the escalated connectivity in message passing techniques (e.g., LBP, TRW) does not result in the growing of their complexity. They show exactly similar complexities of memory and time to those of 4-connected grid-graphs with exceedingly smaller connectivity [14, 15].

3.7. Software package/toolbox

There are a number of software packages/toolboxes available for the implementation, evaluation, and comparison of different energy minimization algorithms, including free/open-source packages (e.g., ADMB, CVXOPT, NLopt, optimx, pyOpt) and commercial toolboxes (e.g., LINGO, Maple,

DOI: 10.1002/cnm

Table II. Examples of existing software package/toolbox for energy minimization.

Software	Brief Description	References
ADMB	A powerful software package for the development of nonlinear optimization frameworks	[211]
CVXOPT	A free software package written in Python for convex optimization	[212]
NLopt	A free/open-source library for nonlinear optimization	[213]
optimx	A free package supporting the optimization of smooth, nonlinear functions with at most box constraints	[214]
pyOpt	A Python-based free package for formulating and solving nonlinear constrained optimization problems	[215]
LINGO	A comprehensive tool for solving linear, nonlinear, integer, and stochastic optimizations	[216]
Maple	A collection of commands for numerically solving optimization problems (e.g., linear, nonlinear, quadratic, continuous, integer, constrained, unconstrained)	[217]
MATLAB	An optimization toolbox consisting of functions for finding parameters that minimize or maximize objectives while satisfying constraints (e.g., linear, nonlinear, quadratic, multiple maxima, multiple minima, non-smooth)	[218]
Mathematica	An optimization package for solving large-scale multivariate constrained and unconstrained, linear and nonlinear, continuous and integer optimizations	[219]

MATLAB, Mathematica). Table II lists some of these popular software. Readers are referred to CRAN [208] and Wikipedia [209, 210] for more information.

4. APPLICATIONS OF ENERGY MINIMIZATION

Energy minimization has been widely utilized in the analysis of medical images. The book edited by Singh et al. [220] explores various earlier techniques based on deformable models and their applications in medical images. In this section, we focus on the applications in medical image segmentation, registration, reconstruction, motion tracking, and compressed sensing.

4.1. Image segmentation

Medical image segmentation is a particularly difficult process owing to a variety of reasons, including i) low signal-to-noise ratio (SNR) in images owing to the inherent limitations of the imaging acquisition technology, ii) image artifacts because of motions from patients or short time for data acquisition, iii) uncertain object boundaries due to the blending of soft tissue with surrounding tissues, and iv) large tissue variability among the population. To tackle the problems, energy minimization techniques have been extensively investigated and popularly utilized in the segmentation of medical data. Examples include corpus callosum [221, 222, 223, 224], basal ganglia [223], ventricle [223] and cortex [225] of the brain, left ventricular (LV) myocardium (endocardium and epicardium) of the heart [221, 222, 179, 225, 27], bone [179], femur [224], vertebrae [224], prostate gland [27, 40], breast cyst [179], and the bounding wall of the liver [225]. Table III lists some example image data for performance evaluation of the energy minimization-based segmentation algorithms. We can see that most of these algorithms can be applied to 2D and 3D medical image data of various imaging modalities (e.g., MRI, CT, and ultrasound) and some of

Prepared using cnmauth.cls
DOI: 10.1002/cnm

Table III. Examples of medical image data used in energy minimization.

Medical Image Data Set	References
2D MRI brain corpus callosum	[221, 222, 223, 224]
2D MRI brain basal ganglia	[223]
2D MRI brain ventricle	[223]
2D cardiac MRI (LV myocardium)	[221, 222, 179, 225, 27]
2D CT bone	[179]
2D MRI femur	[224]
2D ultrasound breast cyst	[179]
2D liver wall	[225]
3D MRI brain cortex	[225]
3D MRI vertebrae	[224]
3D pelvic MRI (prostate gland)	[27, 40]
Multi-dimensional (ND) data	[27]

them can be further generalized to handle multi-dimensional data. These energy minimization-based segmentation techniques can be roughly grouped into several categories discussed as follows.

4.1.1. Graph-based approaches A graph cut is a subgroup of edges separating the object terminal from background terminal, and the cost of each cut is calculated by adding the costs of those edges that the cut splits. According to combinatorial optimization, the segmentation problem in graph cuts is solved by minimizing an energy function formulated by the boundary term and region term [19, 152], where a graph is a group of nodes/vertices that represent the image pixels/voxels, along with edges that connect the node pairs. Every edge in the graph is allocated a weight/cost of zero or positive value. A graph cut is actually a subgroup of those edges, which split the graph nodes into separate groups. The graph cuts techniques have been popularly applied for segmentation in numerous approaches, including image segmentation [226, 227] and volume segmentation [12, 228]. Essa et al. [229, 230] applied graph cuts for the segmentation of intravascular ultrasound (IVUS) images, where the media-adventitia (MA) border is automatically detected without user intervention, based on the minimization of a cost function derived from features like edge/boundary, shape prior, and texture. The authors in [231, 232] incorporated the edge-based and region-based constraints with graph cut and superpixel for interactive segmentation of the MA border in IVUS images and lumen border in optical coherence tomography (OCT) images. In [233], Li et al. validated the 3D surface segmentation in medical volumes employing graph cuts, but this process is computationally expensive. To improve the efficiency, Yuan et al. [234] applied only one layer of over-segmentation regions instead of voxels to segment medical volumes. Similarly, Lombaert et al. [235] reduced the computation time significantly, where a resolution pyramid was utilized to refine the segmentation results in a hierarchical fashion.

In [49], random walker was presented for accomplishing multi-label image segmentation based on a small set of prelabeled pixels. This method is established on a traditional graph and yields segmentation results by separating quantities specified at the graph vertices. Every unseeded pixel is associated with the seed point label. Such a seed point is the one which a random walker could most possibly approach first, starting from the unseeded pixel, provided that it tends to keep away from crossing the boundaries of objects. Later, Sinop and Grady pointed out that random walker and graph cuts are two specific instantiations of a general seeded image segmentation framework [51].

DOI: 10.1002/cnm

To simultaneously extract the inner and outer walls of the lymphatic vessel in confocal microscopy images, Essa et al. [236, 237] developed an approach based on the hidden Markov model (HMM). The segmentation problem is transformed into minimizing the cost of an s-excess graph, where each graph node corresponds to a hidden state and its weight is defined by the emission probability inferred from two separate Gaussian probability distributions for the vessel borders and background. The inter-relation between neighboring nodes is specified with respect to the transmission probability learned from the Baum-Welch algorithm. In order to minimize the number of tracking points, they employed radial basis functions (RBFs) to parameterize the vessel borders. By applying the s-t cut technique, an optimal solution of the s-excess graph can be obtained in polynomial time. Similarly, Jones et al. [238] segmented the lymph vessel wall automatically by combining optimal surface segmentation (OSS) with HMM. The OSS is utilized to carry out a pre-segmentation of the images, which produces the initial state for the subsequent HMM-based segmentation stage. Although weak edges and noise exist in confocal images, a smooth and coherent extraction of both vessel boundaries is attained, leading to a good visualization of 3D vessel structures.

4.1.2. Level set-based approaches Level set-based segmentation [239, 240, 8, 241] involves minimizing the energy-based snakes by calculating the geodesics or minimal distance curves. A curving contour is first placed in a high-dimensional surface as a zero-level set, which is then deformed for the minimization of a standard measurement described in terms of image gradient and curvature. Various level set-based approaches have been proposed in the literature. Lorigo et al. defined the segmentation problem as minimizing an energy over all 3D curves [242] and applied the level-set scheme for hunting a result. This method extends earlier level-set segmentation approaches to higher co-dimensional cases. To segment image data consisting of two or three types of regions that are identifiable using predetermined statistics, Yezzi et al. described a new statistical method for snakes [243], in which the active contours are evolved using both the local and global image information, and the level-set method is utilized in the implementation to enable initial contours to merge and split automatically. Guo and Vemuri [244] embedded the snake pedal into a level-set scheme, allowing the model to manage topological changes without interference. To overcome the difficulties associated with traditional point models for shape representation using a group of marker points, Tsai et al. [26, 27] applied an Eulerian scheme for the description of shapes utilizing the level-set techniques presented in [239]. In [245], Yeo et al. described a variational approach for level set segmentation with statistical shape prior. By employing kernel density estimation, the integrated shape information allows the proposed model to efficiently deal with complex shapes from noisy and occluded images.

In traditional sequential methods, segmentation and interpolation are performed separately in turn. Some approaches first carry out segmentation of the slices and then interpolate a 3D surface from the segmented 2D contours, while other approaches apply interpolation of the slices first to reconstruct a 3D volume, followed by 3D segmentation. Both of them have limitations in processing 3D and 4D sparse medical data. In [246], the authors achieved segmentation and interpolation simultaneously by incorporating them into a RBF-interpolated level set framework, where the flexibility of level set methods, the numerical stability of RBF-interpolated level set segmentation methods, and the interpolation abilities of RBFs are combined. The interpolation exploits the

DOI: 10.1002/cnm

Table IV. Examples of level set implementation in curve evolution.

Level Set Implementation	References
Standard level set [239, 241]	[243, 249, 250]
Coupled level sets	[225]
Dual level sets	[248]
Multiple level sets	[40]
Multi-phase level sets	[252]
Combination of standard and localized level sets [253]	[247]

segmenting surface and its shape information instead of pixel intensities, leading to higher robustness and accuracy. Additionally, the outlined algorithm supports any spatial configurations of 2D slices with arbitrary positions and orientations.

To keep the computational complexity at a reasonable level, various level-set techniques (see Table IV) have been widely used for the implementation of the curve evolution in both parametric models [243, 247, 40, 248] and non-parametric models [249, 250]. Compared with parametric boundary representations, the level-set formulation is independent of parameterization and can cope with topological changes (e.g., splitting and merging of boundaries) during deformation. To enable free initialization of contour/surface and handle more complex topological changes in the image data, Xie and Mirmehdi [251] brought the RBF-interpolated level sets into the region-based active contour model. In comparison with conventional level set-based frameworks, the proposed implicit active model does not require periodic re-initialisation and enables coarser computational grid, potentially benefiting the modelling in high-dimensional spaces.

4.1.3. Deformable model-based approaches Among many energy minimization approaches, deformable models gained popularity in segmenting medical images after Kass et al. introduced the snakes model [1]. In deformable models, a curve or a surface is able to evolve, controlled by the internal constraints preserving the model smoothness throughout the deformation, and the external constraints drawing the models to desired features in the images (e.g., object boundary). The typical internal forces are shape priors, while the external forces mainly include multi-scale Gaussian potential force [1], pressure/balloon force [7, 8, 254, 255], gradient vector flow (GVF) [256, 257], distance potential force [258], dynamic distance force [259], interactive spring and volcano forces [1], curvature vector flow [260], diffused region forces [261], Lorentz force [262], geodesic GVF [263], geometric potential force [264, 265], and magnetostatic forces [248]. Deformable model-based approaches can accommodate to complicated variations in object shapes and integrate shape priors for regularizing segmentations. In the literature, deformable models are referred to with a variety of names, e.g., deformable contours or surfaces, active contours or surfaces, snakes, balloons, etc. Figure. 1 shows several examples of using deformable models for contour/surface extraction.

The representation of object boundaries and utilized external force fields usually vary in different deformable models. Accordingly, previous strategies can be categorized into several groups: gradient-based approaches [257, 263, 248], region-based approaches [267, 250, 268], shape-based approaches [269, 270, 271, 92], and hybrid approaches [261, 272]. Depending on the variety of the representation of the curve/surface evolution during deformation, some techniques can

DOI: 10.1002/cnm

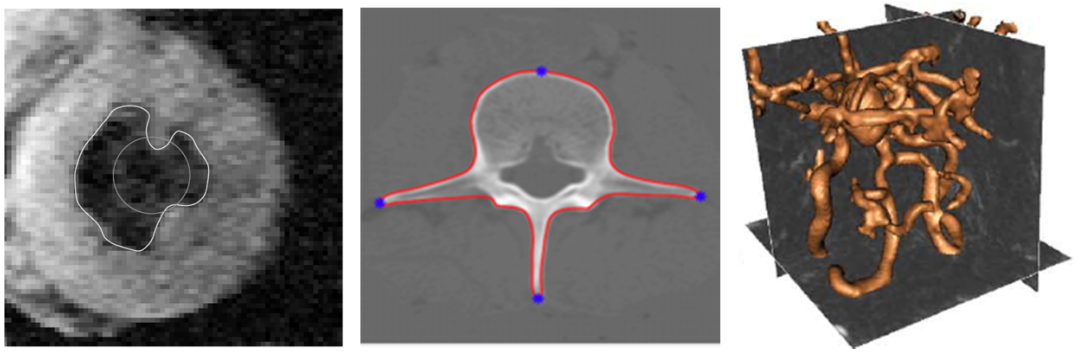

Figure 1. Examples of medical image segmentation using deformable models. (a) 2D heart LV inner wall extraction [256] (circle: initial contour; curve: final result), (b) 2D spinal vertebrae boundary extraction [266], and (c) 3D cerebral artery extraction [265].

be categorized into explicit models or parametric models [1] and implicit models or geometric models [273].

Boundary/gradient-based deformable models The classical snake models [1] are used extensively in medical image segmentation. Starting from an initialized curve near the object, the curve keeps moving towards its interior normal and stops at the desired object boundary. An edge detector relying upon the image gradient features is generally used as the constraint to stop the curve evolution. However, the fundamental issues related to initialization and slow convergence to the concavity of the boundaries hinder the applications of them. In case that they are not placed in the vicinity of the contours, the traditional snakes would not be drawn by them and may straighten to a line. An early solution was developed by Cohen [7], which adds a deflation/inflation force to the contour attempting to shrink/grow it towards edges. In this approach, the convergence does not require the starting curve to be in the neighborhood of the solution. The curve neglects weak edges and stops so long as a strong edge is met.

The geometric snakes [274] and succeeding geodesic snakes [8] (GAC) are both novel schemes for the extraction of object boundaries. The proposed approaches formulated the classical snakes or active contours approach as discovering the geodesics or curves of minimal distance in Riemannian space acquired from the image content. The shortest curves are those along the edges of the image, thus GACs are able to automatically shrink the curves to the edges. This technique enables concurrent identification of internal as well as external boundaries in multiple objects with no special procedure of contour tracking and it does not require special stopping criteria. Yezzi et al. [179] developed a similar method for detecting edges and segmenting medical images on the basis of the image-dependent Riemannian metrics along with corresponding flow of gradients for the active contour models. However, they still have troubles with handling the image noises, weak edges, as well as boundary concavities.

Rather than rely on a constant force, Xu and Prince [257] introduced a novel external force model termed GVF (gradient vector flow) field which extends the influence of edges off into homogeneous image regions, thus expanding the boundary capture scope of snake models. The generalized GVF [256, 257] further improves the performance. However, it has convergence problems, due

DOI: 10.1002/cnm

to the saddle/stationary points in the force field. Later, Xie and Mirmehdi [248, 275] proposed a model based on the edge information: MAC (magnetostatic active contour) model, introducing a fresh external force field constructed from the presumed magnetic force between object boundaries and active contours. The MAC model demonstrates remarkable advancements in handling broken boundary, weak edge as well as complex geometry. The drawback of this approach is that it is difficult to directly apply its magnetostatics-based analogy to high-dimensional images including 3D images. In [265], Yeo et al. presented a deformable model for 3D images where a geometrically induced external force field was introduced, termed geometric potential force (GPF) field. In order to automatically facilitate the topological changes, level-set scheme is utilized to evolve this deformable model. Owing to the bi-directionality of the GPF field, the proposed 3D model can adapt to any cross-boundary initializations. This will help the segmentation of complex geometries and favors the processing of broken boundary and weak edge in images. By strengthening the geometrical interaction field using a non-local algorithm that preserves edges, this model can successfully deal with noise in images. In comparison with the magnetic force field [248], the described vector force field is a generalized extension of the MAC field. However, its expansion to higher-dimensional images is rather straightforward.

The gradient-based approaches have been proved effective in case that the gradients in images are practical indicators of object boundaries and the prior knowledge is limited. However, these techniques need good initialization and they may have difficulties to accomplish invariance of initialization and robustness of convergence, particularly for the segmentation of 3D image objects with complicated shapes and geometric structures. It is worthwhile to note that Xie developed a snakes model employing edge information that is invariant to initializations [275], offering considerable space for contour initialization.

Region-based/edge-free deformable models In the classical active contour models, the stopping term generally depends on the gradient of the image. For objects with blurred boundaries owing to their inherent nature or noises, the gradient-based snakes methods are not applicable. In order to overcome the difficulty, Chan and Vese [89] proposed a new model integrating several algorithms including curve evolution, level-sets, and Mumford-Shah function [82], which is capable of extracting objects of interest in an image with boundaries that are smooth or not necessarily described with respect to gradients. The initialized curve may be at any place in an image and internal contours are extracted automatically. This technique is denoted as ACWE (active contours without edges), which has turned to the canonical example of energy minimization for image segmentation. In [276], Tsai et al. presented a comparable technique to tackle the issue of simultaneous segmentation and smoothing of an image, which addresses the Mumford-Shah paradigm from an aspect of curve evolution. Both methods assume the image consists of regions with intensities that are roughly piecewise constant, and subsequently extract the objects according to the average intensities in both sides of the contour. The methods are able to extract objects with boundaries that are smoothly diversified. However, they may have problems in manipulating image regions with inhomogeneous intensities. In such cases, a piecewise-smooth approximation will be more suitable. Thus, Chan et al. [277] developed an extension to the piecewise-smooth function. In addition, other region-based approaches [243, 247] supposed that the objects in an image are made

DOI: 10.1002/cnm

up of different regional features (e.g. means, variances, textures). In practice, this is generally not the case for real images because of inhomogeneous intensities and the nature of multi-modality.

Shape-based deformable models In case both boundary and region statistics cannot be derived from the image, shape prior information is useful that provides resilience to noisy, missing or diffuse boundaries. A variety of models [278, 269, 270, 224, 279, 271, 280, 27, 281, 282, 283, 92] have been successfully developed to secure the boundary smoothness by means of minimizing the curvature, segmentation area, curve length, or the dissimilarity between the shape of the segmentation and some prior model of the shape.

Unlike previous methods that designed individual shape models specific to particular shapes, the authors in [221] integrated the global shape information into segmentation task, utilizing an elliptic Fourier decomposition of a large number of salient feature points to parameterize the segmenting curve, and then adding Gaussian priors to the coefficients of Fourier decomposition. Cootes et al. [269] designed a parametric point distribution model to describe the segmenting curve by linearly combining the eigenvectors that reflect variations in the mean shapes. To locate an object's boundary, Wang and Staib [223] constructed a statistical model of shape variations based on the correspondence points in a group of training samples. Throughout the optimization process, the parameters of both the pose and shape are searched concurrently, yielding a more desirable optimum. During the searching process, they adopted consecutive penalizing criteria with regard to a Gaussian distribution obtained from a training set for the prior parameters of pose and shape. Moreover, the proposed optimization technique is more efficient and robust than the one presented in [269]. In [284], Duta and Sonka applied an active shape model (ASM) for extracting subcortical structure from MRI brain image data, which integrates an outlier detection scheme in every deformation process. Following that, Frangi et al. introduced a novel framework [285] to produce salient feature points automatically for the construction of 3D statistical shape model, in respect with automatic detection of a compact feature point mesh from an atlas built on the training shape samples. Later on, these extracted salient feature points will be warped to every shape in the training set by means of a non-rigid volumetric registration procedure employing hierarchical B-spline deformation models.

Leventon et al. presented a prior parametric shape model [224], which incorporates prior shape information into GAC model [8] to restrict its flow. The segmenting curve evolves according to the gradient force of the image and the competing force exerted by the estimated shape. The shape parameters are calculated based on the image gradients and the current curve position. In [26, 27], Tsai et al. proposed a method for curve evolution employing the shape information to segment medical images with known object categories. Motivated by Leventon's work [224], they obtained a parametric model to represent the segmenting curve implicitly, where principal component analysis (PCA) is applied to an accumulation of signed distance functions representing the training data. After that, they calculated the parameters of such an implicit representation via gradient descent to make a region-based energy function minimized [89] for segmentation. During the training phase, point correspondence is not required, as the implicit representation is set in an Eulerian framework. This method could cope with topological changes of the segmenting curve, and it is robust to contour initializations and noises. To further improve the performance, the authors in [40] extended the shape-based deformable model [26, 27] to multiple shapes. The new segmentation

DOI: 10.1002/cnm

Table V. Examples of shape representation in energy minimization techniques.

Shape Representation	References
Explicit parametric contour/surface model	[1, 286, 258, 287]
Explicit spring-mass model	[3, 288]
Explicit figural model (M-Reps)	[289]
Explicit topology adaptive contour/surface model	[290, 291, 255, 292]
Elliptic Fourier decomposition of various landmark points	[221]
Parametric point distribution model with variations from the mean shape	[269]
Statistical point models obtained from a training set using PCA of the covariance matrix	[223]
Graph-based shape representation	[146, 293, 19, 152, 12, 13, 49]
3D statistical shape models with a dense mesh of landmark points from a training set	[285]
Prior parametric shape model incorporating shape priors into GAC model	[224]
Tubular spring-mass model (vessel crawlers)	[294]
Implicit parametric representation of a single shape by the signed distance function	[239, 8, 26, 27]
Implicit representation of multiple shape classes by multiple signed distance functions	[40]
Implicit representation of shapes by probability maps	[92]
Implicit multi-class shape representation	[295, 267, 296]
Graph-based multi-class shape representation	[152, 297]

framework enables simultaneous segmentation of multiple shapes by employing multiple signed distance functions implicitly representing the multiple shape types in an image. To obtain a coupling between these shapes and capture the co-variations among them, a parametric model is derived for such a newly proposed representation where PCA is applied to an accumulation of multiple signed distance functions. In order to minimize the cost criterion based on mutual information, this multi-shape model's parameters will be estimated subsequently. The resulting algorithm has the ability of capturing large shape variabilities and it achieves good robustness to additive noise.

For shape-based segmentation algorithms, the representation of various shapes is of great significance. Table V gives some examples that can represent single or multiple shapes with high computational efficiency and accuracy, are effective in acquiring large shape variabilities and are capable of dealing with topological changes.

Hybrid deformable models Gradient-based boundary finding and region-based segmentation are two conventional approaches to medical image segmentation. Both methods have their advantages and limitations. To integrate the strengths of the two approaches, Chakraborty et al. [222, 298] formed a hybrid segmentation model by applying the Green's theorem to obtain the boundary of a homogeneous area with classified regions and combining it with a boundary finder involving gray-level image gradients. Such an integrated method achieves robustness to noises as well as poor boundary initializations. The authors in [299] proposed a unified framework to match an atlas to a target image using intensity as well as shape information. The deformable anatomical atlas can be built using a statistical appearance model constructed from a set of labeled training examples, delineating acceptable variations in intensity and shape in medical images.

In [300], the directions of image gradients are employed for the measurement of alignment, in combination with the criterion of minimal variance [89] and the geodesic active contour (GAC). The

Prepared using cnmauth.cls
DOI: 10.1002/cnm

author applied this alignment measurement to optimize the curve orientation according to image gradients. In association with the geodesic measurement based on image gradients as well as the criterion of minimal variance based on image regions, the alignment measurement is later utilized to attract the contour toward image boundaries. To link the gradients and regional information effectively, the parameters of different measurements need to be tuned carefully. Moreover, in this alignment measurement, the author exclusively adopted the local information of edges without considering the edge information of those pixels with a large distance from the contour.

In [272], Bresson et al. developed a variational model for segmenting objects in medical images, attempting to incorporate the complementary shape, piecewise-constant, and edge terms into one formulation. With many different competing presumptions on the images and objects intrinsic in every formulation, a trade-off of a certain extent should be constructed in the final functional, namely, the weights of different terms ought to be assigned. Otherwise, the weights may result in remarkable errors. McIntosh and Hamarneh [301] proved that optimizing the weights can dramatically reduce the errors in big data sets. Nevertheless, manual optimization of these weights is time-consuming and boring even for one image, without assurance of achieving accurate segmentation results. A variety of recent works [70, 301, 302, 303] attempt to find a way of assigning the weights automatically.

Coupled deformable models In the literature, various coupled snakes models were developed for segmenting the medical image data. For objects like brain cortex with an almost unchanging thickness, they can be modeled in terms of a volumetric layer described by two bounding surfaces along with the in-between homogeneity. Inspired by this truth, Zeng et al. designed a coupled-surface propagation scheme [225], in which both embedded surfaces are evolved concurrently to identify the internal along with external object boundaries in brain cortex images. In this technique, every deforming surface is controlled by a coupling force as well as an information force acquired from the image, preserving a certain distance between the two surfaces. Chan and Vese [252] proposed an efficient multi-phase level-set formulation to segment images based on the piecewise-constant Mumford-Shah model, where only p level-set functions are required for the representation of 2^p image segments. These functions are coupled with each other via an energy functional. To evolve several sets of contours concurrently, the authors in [243, 247] obtained multiple curve evolution equations coupled with each other from only one global energy functional. In this coupled curve evolution model, the evolution of every curve always relies on each image pixel and it is coupled to every other curve evolution straightforwardly, despite the mutual proximity of them. Similarly, Tsai et al. presented a parametric multi-shape model [40] for the segmentation of medical images. The proposed model is capable of obtaining a coupling between multiple shapes and capturing predominant co-variations that are shared in various types of shapes. Since the deformation of a shape always relies on every other shape despite their proximity, the benefit of adopting only one mutual information-based cost criterion is that it can enhance the multi-shape coupling more likely.

Mutual information-based deformable models Mutual information (also known as relative entropy) fundamentally comes from the information theory (refer to [304]), which statistically evaluates the dependence between variable pairs or the information one variable holds for another

DOI: 10.1002/cnm

one. It has been introduced for the segmentation of images recently. In [249, 250], Kim et al. presented an information-theoretic technique integrating curve evolution with non-parametric statistics, where the image segmentation task is formulated as maximizing the mutual information shared between region labels and pixel intensities, susceptible to the constraints on the entire boundary length of regions. The optimization of such an information-theoretic task is solved by utilizing the curve-evolution strategies and acquiring the corresponding flow of gradients. Because of the non-parametric aspect of this formulation, the proposed technique has no limitations on the probability distribution for the image regions or desire the a priori extraction of representative statistics for each region, when compared with many currently available curve evolution-based techniques. The authors in [40] developed a parametric model for multiple shapes that is combined with a framework relying on mutual information for the segmentation of medical images. The parameters of this model are estimated for the minimization of the generalized energy function based on mutual information that was originally suggested in [249] for segmenting images.

The segmentation schemes based on mutual information solely presume that the pixel intensities in individual image regions are statistically dependent with each other, and have no requirement of a specific probability distribution for these regions. In comparison with edge-based methods and those statistical models relying on mean/variance, the mutual information-based models are able to automatically deal with a broader scope of segmentation problems.

Explicit deformable models In explicit deformable models (or parametric deformable models) [1, 7, 221, 287, 305, 306, 70], surfaces or curves are represented with regard to their parametric form explicitly throughout the deformation, which are capable of tracking the points on curves/surfaces over time. This enables the user to interact with the model directly, leading to a representation that can be easily implemented in real time. To locate the object boundary in an image, parametric curves are initialized and then pushed to advance towards the potential energy optima by the constraints of internal and external forces. Numerically, parametric deformable contours can be implemented using the greedy algorithm [141], finite difference method [1], and dynamic programming [307]. On the other hand, parametric deformable surfaces can be implemented by utilizing the finite difference method [308] and finite element method [309, 287]. The finite difference approach is very efficient and it only needs local operations, while the finite element technique is computationally expensive, but it can be easily modified to suit deformable surfaces represented by irregular meshes.

Due to the parameterization of the surfaces or curves, explicit models usually have troubles in coping with topological changes like splitting or merging throughout the deformation. To tackle the difficulties, McInerney and Terzopoulos proposed deformable models adaptive to topologies [255], where the deformable surfaces are formulated in regard to an affine decomposition of cell images so as to process the topological changes which can generally be found in medical images. In this model, a periodical reparameterization scheme is needed for dealing with complicated shapes and topological changes. The presented method can be easily applied for the segmentation of complicated anatomies in medical images, but this model achieves good performance only if it is needed to inflate/deflate everywhere, thus hinder its applications in practice. Besides, there are many other approaches [310, 311] available in the literature for handling topological changes. In these methods, a range of heuristic schemes are usually utilized for extracting self-intersections and

DOI: 10.1002/cnm

manipulating splitting/merging during deformation, which may take much time. Furthermore, they may have a poor performance in case of working on complicated topological structures.

Implicit deformable models In order to overcome the shortcomings of explicit deformable models, implicit models (or geometric deformable models) [274, 240, 273, 8, 179, 244, 26, 300, 27, 263, 40, 261, 265] were proposed, in which surfaces/curves are implicitly depicted by a level-set embedded in a high-dimensional scalar function, based on the prevailing level-set technique [241] and the theory of curve evolution [312, 313]. In implicit deformable models, the curves/surfaces evolve according to geometric measures (e.g., curvature, unit normal). The parameterizations of them are independent of the evolution, thus adaptivity to topological changes can be accommodated automatically. The numerical implementation of implicit deformable models can be achieved by using the methods described in [76, 273, 241]. These methods have been popularly utilized for segmenting anatomical structures in volumetric medical images [314, 27, 40, 315, 316].

Numerous extensions have been proposed to the traditional parametric and geometric deformable models by incorporating additional prior knowledge together with global shape properties into the models, leading to more robust and accurate results. Especially, the integration of global shape properties can remarkably increase the robustness of initialization. In [221], Staib and Duncan presented a Fourier representation for the parameterization of deformable surfaces and contours. By truncating the series, they acquired a compact representation of smooth shapes. For the characterization of global shape property, the authors attained a geometric shape description as well. In this technique, a Bayesian approach is employed for incorporating the prior information, and the maximization of the posterior probability function is accomplished utilizing an algorithm of gradient ascent, which can be secured through other means (e.g., genetic algorithm [317]). Terzopoulos and Metaxas [309] proposed a deformable superquadric approach which combines global as well as local shape features. This deformable superquadric model has been extended to multi-resolution shape modeling [318] and multi-level shape representation [319]. In [318], Vemuri and Radisavljevic exploited the multi-resolution wavelet decomposition to parameterize local deformations, which provides a powerful transformation from local descriptors to global shape deformations. In [319], Metaxas et al. achieved a hierarchical shape description for dense along with sparse range data, utilizing the locally adaptive finite element technique and global deformation of the dynamic model. To integrate the prior information of shapes, Cootes et al. developed the ASM models [320, 269], where the prior models are based on a set of aligned landmark points manually labeled on the training images, instead of the parameterization. To describe the variability in the coordinates of every point as a prior model, they designed the point distribution model. The model is employed to restrict the action of deformable models [284, 223], which exclusively allows deformations producing shapes indistinguishable to those of the training examples. As the prior models of ASM do not contain intensity variation across images, Cootes et al. extended the ASM to active appearance models (AAM) [321, 299]. The prior model of AAM represents shape as well as texture variations of the objects across images, leading to more robust results. In [244], Guo and Vemuri proposed extensions to a geometric deformable model by integrating the global shape prior information. Like traditional geometric models, this hybrid model can automatically manage topological changes. The authors in [322] presented a robust topology independent solution for segmenting the left ventricle and right ventricle in cardiac images utilizing an implicit deformable

DOI: 10.1002/cnm

model that is constrained by the conditional random fields (CRFs). By employing a graphical model, this deformable model is successfully incorporated with the CRFs strategy. The model deformation is completed by estimating a maximum a posteriori (MAP) problem, in which the desired conditional probability is divided into two terms: the internal one and the one driven by images. In this framework, both patch- and pixel-based discriminative CRFs are applied to attain smooth probability fields with respect to the corresponding image features (i.e., the pixel intensity and the multi-scale texture). Moreover, the variations of local features are handled by refreshing the model interior statistics.

4.1.4. Atlas-based approaches In atlas-based medical image segmentation techniques, some specific anatomical structures of interest are first labeled by an expert from images collected by different imaging modalities like MRI and CT. The manually labeled diagram is called atlas, which contains the prior knowledge and captures the anatomical variations in the medical data. The relevant anatomies are then extracted based on the atlas, which can help the clinicians to diagnose diseases. These atlas-based segmentation methods have been popularly utilized in numerous clinical tasks, e.g., kidney and renal cortex extraction for perfusion measurement, coronary artery segmentation for tracking coronary vessels, brain structure segmentation for disease diagnosis, and heart detection for cardiac analysis [323, 324, 325].

Image alignment/registration plays a crucial role in atlas-based segmentation, which finds a mapping between a target image to be segmented and a reference image by minimizing a predefined energy function. A variety of algorithms have been presented in the literature, including affine/linear registration [326] and nonlinear registration [327]. Mostly, these atlas-based techniques label the target structures in an automated fashion. Han et al. [324] designed an automatic atlas-based algorithm to segment the head and neck data, utilizing a hierarchical registration method that incorporates both linear and nonlinear registrations. Bauer et al. [328] developed a non-parametric framework for the identification of normal tissues in 3D brain tumor image data, relying upon atlas segmentation that combines non-rigid atlas registration with a tumor growth model. In [329], Fortunati et al. presented an atlas-based approach to automatically extract thermo-sensitive tissues from the CT head and neck images. Taking the large variation in such anatomies into account, they integrated atlas-based registration with intensity-based classification. Graph cuts are utilized to minimize a specified cost function globally.

To further increase the segmentation accuracy, multi-atlas-based segmentation was introduced, where multiple atlases are involved. Klein and Hirsch automatically identified the large cortical structures in MRI brain data combining multiple atlases [330]. The authors in [331] proposed an automatic algorithm for the segmentation of brain anatomies. They combined different atlas segmentations employing a fusion strategy at the decision level, in which each segmentation is propagated to another one according to the estimated anatomical correspondence between an atlas image and a target image. Later, Lijn et al. described an automated scheme to accurately segment the hippocampal volume (hippocampus) from the MRI brain images [332]. The atlas-based registration is incorporated into the voxel-based classification. By employing graph cuts, a cost function is minimized globally, which is made up of both intensity energy and prior energy including spatial prior as well as regularity prior terms. The presented approach achieved a better performance than the multi-atlas-based segmentation techniques.

Prepared using cnmauth.cls
DOI: 10.1002/cnm

Table VI. Examples of measurements for the performance evaluation of energy minimization algorithms.

Performance Metric	References
Average Euclidean distance	[221, 299]
Squared Euclidean distance	[222]
Boundary maximum error, boundary average error, and correspondence average error	[223]
Undirected partial Hausdorff distance	[337, 224]

To enhance the segmentation performance, atlases yielding good results can be selected according to some criteria like minimum Hausdorff distance [333], and multiple atlas results can be weighted and combined using some fusion strategies such as majority voting [334, 335, 336].

4.1.5. Performance evaluation As illustrated in Table VI, various measures have been used to assess/validate the similarity/agreement between the segmented boundary/contour and the ground truth. Most of them are based on the average distance between the corresponding point pairs on two cruves/surfaces. Sometimes, the goodness of a segmentation algorithm is subjectively and qualitatively rated by an expert (e.g., radiologist) [27].

4.2. Image registration

One of the fundamental problems in the medical image analysis field is the registration (also known as alignment, fusion, or matching) of the volumetric images captured by the same imaging modality or at different time points, due to the continuous movement of the anatomical tissue undergoing examination. On the other hand, multi-modal registration of medical images plays a crucial role in clinical practice by integrating the complementary information from multiple images of the same subject that are collected by various imaging modalities. For instance, by registering the MRI images in neurosurgery with the CT images in the established stereotaxy technology, it becomes possible to transfer the tumor coordinates from MRI images to CT stereotaxy. Similarly, for the analysis of brain functions, PET image data yield functional information, whereas MRI images contain anatomical information, thus it is helpful to correlate functional information from PET with MRI for anatomical reference.

Image registration aims to recover the geometric relationship between corresponding points in multiple images. Various schemes have been implemented to tackle the medical image registration problems. The authors in [338] discussed and classified the earlier approaches concerning the registration of multi-modal images. Following that, Maintz and Viergever presented a review on medical image registration techniques covering the literature between 1993 and 1998 [339]. Later in 2003, Pluim et al. provided an overview on mutual information-based registrations [340], and Maes et al. gave a survey on the registration criteria of maximizing the mutual information and their applications for 3D medical image registration [341]. All of these three papers provide a valuable collection of registration algorithms for both mono-modal and multi-modal images.

4.2.1. Mono-modality image registration The medical image registration problem can be modeled as an energy minimization framework. An earlier approach was proposed by Bookstein [342] for the registration of medical data captured using the same imaging modality, employing

Prepared using cnmauth.cls
DOI: 10.1002/cnm

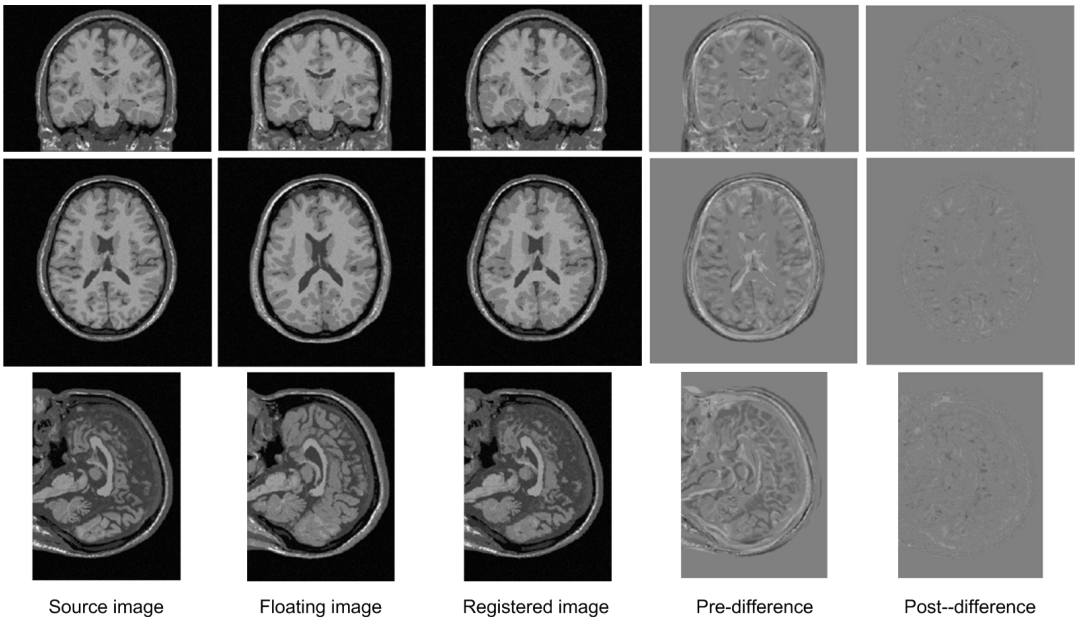

Figure 2. Examples of 3D mono-modality image registration [347].

the decomposition of deformations through principal warps. Later, Gueziec and Ayache [343] developed an efficient technique based on geometric hashing for rigid registration of 3D curves extracted from CT vertebra and head images. A locally affine deformation algorithm for non-rigid registration of two surfaces was presented by Feldmar and Ayache [344], which has no requirement of a prior parameterization or the topological information of the surfaces. In [345], Lavallée and Szeliski introduced a 3D distance map depicted by an octree spline for efficient and accurate registration of 3D anatomical surfaces with a set of 2D projections. Rueckert et al. described a non-rigid approach to align 3D MRI breast images [346]. By interpreting the local and global breast motions with freeform deformations and affine transformations respectively, they obtained a completely automatic registration based on energy minimization. As a result, the presented non-rigid registration method sufficiently reduces the motion artifact in many breast image pairs and it outperforms other rigid/affine registration schemes for matching volumetric breast images. Figure. 2 shows a few examples of 3D registration results across two MRI brain volumes.

To embed the utilization of explicit a prior information into a deformable model, Bajcsy and Kovacic developed a multi-resolution elastic registration algorithm [348] for 3D CT brain data. Following a preliminary global registration, the brain atlas evolves gradually in a hierarchical fashion, which increases the local similarity as well as global coherence. This method presumes that normal brains, minimally at a certain degree of description, contain identical topological structures, but different shape features. This elastically deformable atlas approach implemented to brain image registration can be found in [349, 350, 351, 352, 353]. However, it is susceptible to atlas initializations, i.e., in case that the bias of preliminary rigid registration is rather big, the subsequent elastic matching could be poor. In addition, the existence of neighboring features can possibly make the atlas warp to inaccurate boundaries, and the atlas may not be able to converge to complex object boundaries if no user intervention. To overcome the difficulties, Sandor and Leahy [354] applied the 3D edge detection together with morphological operations to preprocess the brain images, so as

Prepared using cnmauth.cls
DOI: 10.1002/cnm

to produce a smooth delineation of the brain surface based on which the 3D B-spline deformable model may achieve a rapid convergence.

In [347], Tang and Chung formulated the non-rigid medical image alignment as an energy minimization problem, and then achieved its optimization employing the graph-cuts technique through α-expansions. The problem is that the dissimilarity measurement adopted in the constructed cost function is confined to the summation of squared differences and the total amount of absolute differences. This graph-cuts-based approach was extended in [355] by adopting an improved dissimilarity measurement like mutual information. As a result, the registration errors are largely reduced. The non-rigid medical image registration may also be described as a MAP inference problem in a MRF [356, 357, 358], and its optimization can be secured utilizing the pairwise Gibbs energy minimization framework.

To obtain a denser set of point correspondences from a few manually labeled landmark correspondences, the authors in [359] proposed a reliable semi-automatic approach for the registration of 3D aortic meshes. They employed a two-stage process to estimate the local non-rigid transformation based on B-spline freeform deformation, in which the B-spline parameters are optimized through gradient descent method. While lowering the possibility of edge overlapping, this strategy prevents sharp peaks/troughs in the mesh surface and alleviates the over-fitting problem.

As a measurement of the amount of information that one image holds for the other, mutual information has been commonly used for automatic rigid as well as affine mono-modality/intra-modality image registration. Viola and Wells [360, 361] first introduced an efficient and reliable information-theoretic strategy for registering the volumetric medical image data collected using the same imaging modality. Obtained from information theory, the registration is achieved by maximizing the mutual information between the images. Without requiring preprocessing (e.g., edge detection) or prior segmentation, this intensity-based technique attains good performance in fields where edge/gradient-based approaches perform poorly, and it secures more robustness and higher flexibility compared to other schemes based on intensity (e.g., conventional correlation).

4.2.2. Multi-modality image registration Taking advantage of the complementary information between images from different modalities, many techniques employing deformable models have been developed for multi-modality/inter-modality image registration (see Figure. 3). In [362], Moshfeghi developed an elastic matching scheme using contour knowledge for the registration of 2D images acquired by different imaging modalities. This technique was extended by Moshfeghi et al. [363] for 3D multi-modality volume registration using surface knowledge. In their work, a 2D image or 3D volume is formulated as being deformed in respect with another target image or volume. The correspondences between contours/surfaces in two images/volumes are adopted for stretching or warping the evolved image/volume towards its target. Such a procedure is iterated till good registration result is secured.

Guimond et al. [364] described an approach for elastic registration of 3D images of multi-modality, using an iterative strategy that repeats many times between obtaining the coefficients corresponding to polynomial intensity transformations and matching the images utilizing the demons scheme [365]. It supposes a maximum of two function dependencies within anatomical structure intensities existing in those images to align, which hinders its implementations in cases where noise and inhomogeneity exist. To overcome this difficulty, Arce-Santana and Alba [358]

DOI: 10.1002/cnm

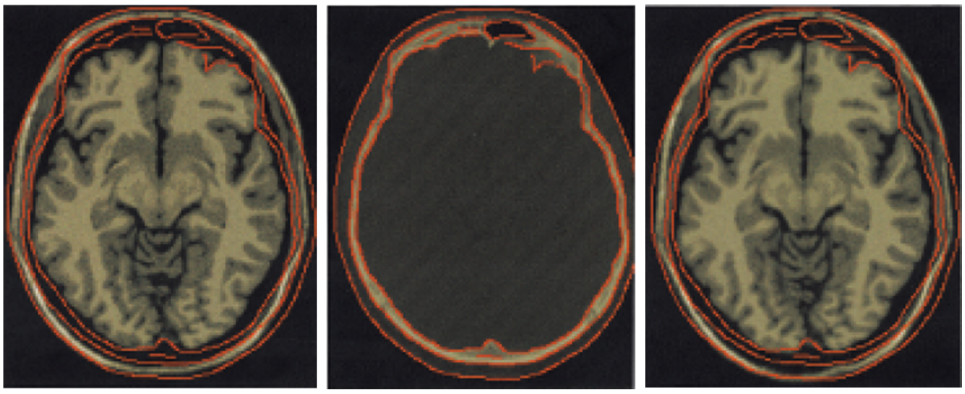

Figure 3. Examples of multi-modality image registration (MRI-CT) [364]. Contours are obtained utilizing a Canny-Deriche edge detector for the CT image, and then overlaid on the other two images for a better evaluation of the registration quality.

presented an intensity-based algorithm relying on Bayesian estimation for 2D multi-modal image registration, where two different types of MRFs are formulated, describing the coefficients related to local polynomial intensity transformations (MRCF) and the parameters corresponding to local geometric transformations (MRGTF), respectively. In this approach, the resulting posteriori energy function is minimized by the Newtonian gradient descent method. Additionally, the proposed registration method achieves great robustness to noise and large inhomogeneity in the images, because of the spatial attributes of MRGTF and MRCF. In [356], Glocker et al. introduced a procedure for deformable image alignment, which links the continuous deformation with the most suitable discrete optimization. Without requiring estimating the derivatives of the adopted energy function, they modeled the registration problem utilizing a discrete MRF objective function and recovered the most proper solution to the specified objective function by means of efficient linear programming applying the primal-dual technique.

Mutual information has also been widely used for automatic rigid and affine multi-modality image registration. The authors in [366, 367, 368, 369, 370, 371] proposed approaches for aligning images of different modalities from the same underlying anatomy (e.g., CT, MRI, and PET brain images), based on the maximization of the information which an image holds for the other one. This necessitates no a priori model of the connection within different imaging modalities, and is robust to the partial occlusions of one of the imagery. In [372], Studholme et al. studied the robustness of standard and normalized mutual information for automatic alignment of 3D medical images of multi-modality, showing that the normalized mutual information outperforms the standard mutual information. To further improve the registration accuracy and robustness, the authors in [373, 374] incorporated the spatial information by extending the intensity-based mutual information with a term based on other derived features such as edges or gradients.

The maximization of mutual information can be achieved using various strategies. In [360, 369, 361], the maximization is attained by locally seeking an optimum via approximating the gradient descent stochastically. Collignon et al. [366] and Maes et al. [367, 370] applied Powell's method to maximize the mutual information measure based on the joint histogram of image intensities obtained from the overlaying area of both images. Studholme et al. employed a multi-resolution

DOI: 10.1002/cnm

scheme [368] to accelerate a heuristic searching strategy for the maximization. They updated the parameters one by one with slight amount in an iterative way. Meyer et al. utilized a downhill simplex searching mechanism for optimizing the parameters of a thin-plate spline-warped or affine alignment [371]. In [375], Positano et al. applied the downhill simplex searching technique and Powell's method for aligning image pairs of identical modality, while genetic algorithm (GA) is employed as an optimization technique in [376] to achieve correctly aligned images of both the same and different modalities. Comparative studies were conducted by Maes at al. [377] to systematically explore the performance of diversified optimization techniques as well as multi-resolution schemes for registration of high-resolution medical images of multi-modality by maximizing the mutual information. By applying a coarse-to-fine multi-resolution technique, they evaluated multiple different nongradient-based together with gradient-based methods including downhill simplex, Powell, steepest gradient descent, Levenberg-Marquardt, conjugate gradient, and quasi-Newton algorithms. The outcome of this evaluation clearly show that in comparison with Powell's approach which is utilized at full image resolution, the Levenberg-Marquardt, conjugate gradient, and simplex methods accomplish an average three-time increment in efficiency employing a two-stage multi-resolution framework, with comparable preciseness as well as equivalent robustness. Similarly, to make the number of local maxima existed in the mutual information-based cost function smaller and make the optimization procedure faster for multi-modality image registration, Zibaeifard and Rahmati [378] proposed a two-level searching scheme where Powell's method is applied in the first stage and a simulated annealing algorithm is exploited in the second stage when the original scope of the searching space is secured for any stochastic searching approach. The described simulated annealing framework is adaptive to normalized mutual information function for handling multiple parameters with respective preliminary scopes.

In addition, mutual information can be applied for non-rigid registration of medical images as well. Palos et al. [379] presented a non-rigid method for the registration of MRI and CT images. In [380], Klein et al. compared the capability of multiple optimization techniques for non-rigid alignment of medical image data by maximizing the mutual information, associated with a deformation field described in terms of cubic B-spline. The experiments involving various organs and imaging modalities indicate that the Robbins-Monro strategy achieves highest efficiency, while the nonlinear conjugate gradient and quasi-Newton schemes yield a moderately better preciseness compared to the Robbins-Monro strategy, at the cost of much more computational time.

The value of mutual information may be influenced by interpolation, which can potentially restrict the image registration quality. Freire and Godinho [381] conducted a cross-validation study to evaluate how the interpolation level and the elaboration of joint histogram after voxel intensity rescaling influence the overall accuracy of mutual information-based registration of both mono-modality and multi-modality images. They concluded that cubic-spline interpolation associated with multiple joint histogram update is the best. Following that, Tsao [382] investigated the interpolation artifacts from eight interpolators for multi-modality image alignment and the strategies of high efficiency to reduce the artifacts. This work showed that several strategies could be adopted to assure the registration accuracy and enhance the registration robustness, namely, nearest neighbor with a histogram blurring, preventing a great intensity bin number, nearest neighbor with a jittered sampling, as well as resampling the image data in a rolling direction employing unequal sizes of voxels.

DOI: 10.1002/cnm

Table VII. Examples of energy minimization-based registration algorithms for medical images.

Classification Criterion	References
2D image data	[362, 353, 368, 375, 382, 376, 347, 358, 355]
3D image data	[348, 350, 344, 343, 363, 351, 345, 360, 367, 368, 369, 370, 371, 361, 377, 372, 381, 373, 364, 378, 376, 347, 355]
MRI image data	[350, 344, 351, 353, 367, 368, 369, 370, 371, 377, 372, 381, 373, 364, 375, 382, 378, 376, 347, 358, 355]
CT image data	[348, 343, 345, 360, 368, 361, 367, 369, 370, 371, 377, 372, 373, 364, 378, 376, 358]
PET image data	[367, 369, 370, 371, 372, 381, 373, 378]
SPECT image data	[371, 382]
Rigid registration	[343, 366, 360, 367, 368, 369, 370, 371, 361, 372, 373, 374]
Non-rigid registration	[344, 379, 380, 347, 356, 357, 358]
Mono-modality registration	[348, 342, 349, 350, 344, 343, 351, 345, 354, 352, 360, 353, 361, 347, 356, 357, 358, 355]
Multi-modality registration	[362, 363, 366, 367, 368, 369, 370, 371, 372, 373, 364, 379, 378, 376, 380, 356, 358, 374]

The quality of those registration algorithms which rely upon aligning object surfaces or anatomical landmark points is susceptible to the errors during segmentation, while the voxel-based approaches like mutual information measures the statistical dependence between the intensities associated with every voxel pair within the two images, supposing that it is maximized in case that both images will be matched geometrically. Maximizing the mutual information is a rather strong and generic principle, permitting completely automatic, robust as well as accurate alignment of multi-modal image data of dissimilar contrast and resolution without requiring preprocessing, prior segmentation, or user intervention, making them well suited for a variety of applications in clinical practice.

A simple quality measurement for the image registration of the same modality is the total squared difference of all the voxel pairs. However, it is not effective for the measurement of the registration of multi-modal images. Table VII gives a summary of the registration methods discussed above, from which we can see that most of these techniques are well suited for the alignment of the brain and the LV of the heart data, and some earlier approaches are relevant to wrist, vertebra and skull image fusion as well. The imaging modality involved are mainly MRI, CT and PET. Some algorithms are applicable for 2D medical image registration and some are valid on 3D image registration, while others are appropriate for aligning 2D along with 3D images. Both rigid and non-rigid image alignment strategies are available in the field of medical image analysis. Some of them are for the registration of images of the same modality, while some others are related to images of different modalities. It is worth noting that some energy minimization techniques can be applied to both mono- and multi-modality medical image registration [381, 373, 376, 356, 358].

4.3. Image reconstruction

In medical image analysis, finding a smooth surface which fits a range of specific unstructured 3D data plays a crucial role. Generally, deformable model-based approaches present satisfactory results. The superiority of parametric deformable models (e.g., superquadrics [309]) is their ability of describing shapes using a few parameters. However, it may be difficult for them to accurately calculate complicated surfaces using a limited range of shapes represented using superquadrics. In order to elaborate the calculation, Bardinet et al. [383] introduced a superquadric-fitting-based

Prepared using cnmauth.cls
DOI: 10.1002/cnm

parametric deformable model, which is succeeded by freeform deformations to suit a range of unstructured 3D data. By applying superellipsoids as the initial shapes, this model has been successfully used for reconstructing the left ventricle (LV) from many different 3D cardiac images. The significant benefit of such a hybrid scheme is that a complicated shape could be depicted compactly and the fitting is very robust, even though outliers or sparse data are presented. The authors in [384] developed a bank of graduated nonconvexity (GNC) techniques for minimizing the MAP energies comprising terms of convex data-fidelity along with nonconvex/nonsmooth potential functions, which can be utilized for reconstructing the emission computed tomography (ECT) data. The proposed algorithm is of high efficiency and succeeds in dealing with the ill-posed problems associated with the traditional iterated conditional modes (ICMs) method [75]. Raj et al. [385] applied the MRF for reconstructing the MRI images by solving the resultant energy minimization problem via graph cuts, greatly improving the reconstruction performance in terms of the SNR. In [386], Benameur et al. considered the 3D reconstruction of scoliotic ribcages as a group of optimization issues, each of which is accompanied and controlled by a particular kind of pathological deformation discovered from a typical training population of scoliotic ribcages and effectively identified employing the probabilistic PCA mixture model. The presented method can efficiently minimizes the X-ray dose and it can be exploited to many other reconstruction tasks of medical data such as knee and pelvis, where an anatomical structure data set is obtained involving at least two radiographic projections.

Generally, the motion estimation and image reconstruction problems are treated separately. For example, Cunningham et al. [387] specified the 3D reconstruction of first-pass low-count cardiac single-photon emission computed tomography (SPECT) images in a Bayesian scheme by employing the geometric deformable models. They minimized the whole surface curvature using a weighted prior. Lalush and Tsui [388] presented an effective block-iterative MAP technique with Gibbs priors for 4D reconstruction of gated cardiac perfusion SPECT image data. Based on the minimization of an objective function by a two-step iterative procedure, Mair et al. [389] investigated the simultaneous estimation of non-rigid motion vector field as well as emission medical image data during the dynamic imaging processes like gated cardiac ECT. The gated data are subsequently utilized for reconstructing the myocardial motion and emission intensity.

Because of the noise in medical images and distortions in staining, it is challengeable to reconstruct the 3D models from a group of nanoscale images. In [390], Yang and Choe proposed a method to automatically obtain the 2D contours and track them via cross-sections to construct 3D structures of the electron microscopy (EM) image stack, where graph-cuts algorithm is employed to achieve a globally optimal result of a cost function comprising the distance function, image gray-scale intensity, as well as flux of the gradient vector field. The incorporation of a distance function tackles the ambiguity problems existing in the EM image boundaries. To secure an entire 3D volume structure that includes branching of the EM images, they combined both 3D volumes acquired by implementing the minimization of MAP-MRF cost function utilizing backwards and forwards algorithms of graph-cuts throughout the image data [391]. In this approach, the over-segmented regions using a marker-controlled watershed method are viewed as the graph nodes while the weight of every edge between the node pairs in neighboring image slices are specified according to the information in overlapped areas.

DOI: 10.1002/cnm

4.4. Motion tracking

Energy minimization techniques are also well suited to motion estimation/tracking in medical image data. Among many other approaches, deformable models are popularly applied.

4.4.1. Deformable model-based approaches Metaxas [392] as well as Blake and Isard [393] introduced a large variety of earlier deformable approaches applied in motion tracking. In medical image analysis, deformable models are mainly exploited for the motion tracking by measuring the dynamic heart behavior, particularly the LV dynamics. A simple way is to apply 2D deformable contour models for the segmentation of the LV boundary in every volumetric image and then deformed the contours for the extraction of an updated range of the LV boundaries [394, 395, 396]. With the purpose of tracking non-rigid LV motions in 4D cardiac CT data, McInerney and Terzopoulos implemented this temporal propagation method using a 3D dynamic deformable balloon model [287].

Deformable models [397, 398, 399, 400, 401, 402, 403] have also been applied to magnetic resonance tagged image data that can capture complicated non-rigid motions like twisting of the heart. Kumar and Goldgof [397] applied 2D active contour models to automatically track the spatial modulation of magnetization (SPAMM) grid points in cardiac MRI image data. Together with the thin-plate spline model, the correspondences between grid points are utilized for building a mapping between images, which is then used for motion estimation. In [400, 402, 403], Park et al. exploited a novel group of volumetric physics-based deformable models with parametric global deformations to analyze the 3D LV motions from the MRI-SPAMM image data. The model parameters are functions and could acquire the local LV shape variation (e.g., contraction, twisting). Furthermore, to obtain a more precise estimation of the LV motions, they constructed a 3D finite-element model with datapoint samples taken from surfaces of the mid-wall, epicardium, and endocardium. Earlier 3D finite-element schemes constructed from boundary delineations of the surfaces of epicardium as well as endocardium for tracking the 3D LV motions from the MRI-SPAMM data can be found in [404, 405, 406, 401]. O'Donnell et al. introduced a deformable thick-walled ellipsoid model [407] with local and global elements of 3D cardiac motions, resembling the LV of the heart. The deformation parameters are adapted to a priori stripe and contour data. They applied a model-based smoothing constraint for the penalization of the deformation variations and the interpolation of the scattered guide data, which can interactively produce a global smooth deformation intuitively without separating global from local deformations. Aiming to locate and track the deformations of the LV in the SPAMM MRI data, Radeva et al. proposed a 3D tensor product B-spline model [408], where the parameters are directly adapted to the stripes in the images of short-axis applying a snakes-like approach. In this work, a volumetric model is employed as two coupled B-spline surfaces of epicardium and endocardium for tracking the motions from image boundaries and confining the image constraints to myocardium only. By exploring a B-solid for the tagged images, the information contained in various slice images is adopted, which makes the proposed method achieve noise robustness. To track the motions of the myocardium in the LV, the authors in [409] developed a completely automatic approach according to the non-rigid image registration utilizing multi-level deformations of free form. The heart images of short-axis and long-axis are both involved for the estimation of the complete 4D motion field within the myocardium.

Prepared using cnmauth.cls
DOI: 10.1002/cnm

Compared to other techniques, the localization of tags and reconstruction of deformation fields are fulfilled concurrently in this method. Similarly, non-rigid image registration was applied to track the myocardial motion involving untagged as well as 3D tagged MRI images from short-axis and long-axis views [410], where a series of tagged and untagged MRI images acquired in the cardiac process are registered with a group of reference images collected in the course of end-diastole. A key benefit is the parallel adoption of supplementary motion knowledge embraced in tagged as well as untagged image data.

Unlike previous approaches, Young [411] proposed a method for directly tracking the motions of 3D heart wall from tagged MRI images of short-axis and long-axis, without requiring the prior detection of both outer and inner ventricle boundaries or locations of tag stripes. Similarly, a finite-element scheme was utilized for depicting the heart motions and heart shapes. The author attained an objective function for measuring the matching similarity between image stripes and model tags, and minimized it by allowing the model to evolve immediately which is controlled by the image forces, employing an effective algorithm to calculate the motion constraints obtained from images. In [412], Chen et al. applied a meshless deformable model that incorporates smoothness, phase, and image knowledge to refine the tracking of interior myocardial motions from a tagged MRI acquisition. By employing the displacement magnitude as the smoothness constraint, they improved the overall motion tracking accuracy. In addition, the proposed tracking method can increase the credibility and throughput of quantitative cardiac analysis.

For the tracking of 3D living tissues in minimally invasive surgery, Wong et al. [413] developed an innovative deformable model on the basis of a quasi-spherical triangle. They utilized the triangle's three vertices for the description of the affine model and integrated a curving parameter to fit the nonlinear properties of soft-tissue motions. These parameters can be estimated using an effective second-order minimization approach [414]. To deal with the occlusion problems, they proposed a motion prediction framework by investigating the motion signals' peak-valley features.

4.4.2. Modal analysis-based approaches In [415], Pentland and Horowitz introduced a dynamic deformable model on the basis of modal analysis using the finite-element technique to estimate the rigid and non-rigid motions of the LV. However, it may be difficult to observe some kind of rigid-body motion and non-rigid motion in case of orthographic projections. Moreover, the adoption of optical flow data instead of feature points necessitates the integration of past object motions for determining the present shape along with position of the object. Similarly, Nastar and Ayache [416, 417] developed a physically-based deformable model for tracking and analyzing non-rigid dynamic structure motions in 2D/3D medical data. They applied modal analysis of the motion for an enhanced fitting of the physical model, which enables a dense depiction of a complicated deformation using less parameters. Later, they extended this work to non-rigid motion analysis of spatio-temporal (4D) medical images by linking modal analysis with Fourier analysis [418]. In this frequency-based framework, the authors utilized a physically-based deformable surface to approximate the non-rigid motions of the evolving object contours dynamically. To decrease the deformation parameter number, they applied the modal analysis that yields a spatial surface smoothing. On the other hand, the Fourier analysis of the temporal part of the principal deformation spectrum produces a useful data smoothing in the time domain. Thus, a complicated non-rigid

DOI: 10.1002/cnm

deformation process can be expressed by only several coefficients including the major Fourier harmonics and the predominant modal amplitudes.

4.4.3. Deformable superquadric-based approaches As described in [419], the deformable superquadrics [309] can be used for estimating non-rigid motions from insufficient data with time-variance and noise-corruption. Chen et al. presented a hierarchical decomposition-based approach [420] for estimating the global rigid LV motions from angiographic data. The hierarchical motion model is assembled by integrating a locally deformable surface secured through making spherical harmonic primitives of shape modeling with a globally deformable superquadric, which yields a coarse-to-fine decomposition that hierarchically characterizes the LV motions, and results in fairly effective estimation techniques. In [421], Bardinet et al. proposed a compact and robust deformable superquadric model to track anatomical structures of the LV of the heart in 4D image data.

4.4.4. Bending energy approaches The authors in [422, 423, 424] proposed a unified scheme to track the non-rigid LV motions of the endocardial wall from a 3D image sequence, based on bending energy along with surface curvature. By selecting geometrically salient feature points representing the surfaces of epicardium and endocardium respectively, they created a pair of scattered subgroups of surface points. Following that, they modeled all the surface patches around these points as thin flexible plates. By assuming the local and slight evolution of every surface patch during a short period of time, for every salient feature point on the first surface, the authors constructed a hunting window on the surface of the LV in the subsequent time interval. The point correspondence between the two surfaces are found by minimizing the bending energy. Based on the preliminary motion vectors produced by such a matching procedure, they performed a smoothing process to create a compact field of motion vectors on the surfaces of the LV. To improve this bending energy technique in tracking the heart structure from 2D ultrasound image sequences, Cohen et al. [425] incorporated a new energy term which inclines to keep the alignment of points with high curvatures, while securing a displacement vector field with good smoothness in all places.

4.4.5. Curvature-based approaches A confined type of non-rigid motion of the LV is termed as conformal motion, defined as a motion maintaining angles instead of distances between curve pairs on the surface. In conformal motions, the stretching of a surface is equivalent in every direction, but unequal at individual point. To deal with surfaces undertaking conformal motion, a number of approaches have been developed with regard to variations in Gaussian curvatures [426, 427, 428, 429].

4.5. Compressed sensing

Compressed sensing (CS, also called compressive sampling/sensing) [430, 431] is a newly developed numerical paradigm for rapid data acquisition and reconstruction. Compared to traditional methods, it allows comparatively fewer measurements of the image to be collected by means of a linear way (undersampling) while enabling accurate reconstruction through a nonlinear recovery procedure. CS effectively utilizes structured prior information in a generic style of sparsity, i.e., representing an image using several coefficients in an appropriate transform basis (e.g.,

DOI: 10.1002/cnm

wavelet, Fourier). If the image is adequately sparse or compressible, it can be completely recovered from undersampled *k*-space measurements via minimizing a convex function. For the theory or mathematics of CS, we refer the readers to several overviews in the literature [432, 433, 431].

CS has significantly expedited some specific medical image acquisition process like MRI [434, 435, 436, 437], owing to the coded nature of MRI image acquisition along with the transform sparsity of MRI data. In addition, to lower the memory burden as well as the computational complexity, block-based CS [438] is popularly applied in high-dimensional image reconstruction, where an image is divided into non-overlapping parts of small size that are obtained separately but reconstructed collaboratively. Furthermore, Razzaq et al. [439] reported an extension of the CS: locally sparsified CS, in which local sparsity constraints for independent subregions are adopted instead of global constraints. The proposed method may be implemented for reducing the noise artifacts as well as improving the image reconstruction quality without raising the overall sampling rate or slowing down the acquisition procedure.

During the CS-based image reconstruction process, energy minimization techniques are employed to optimize the unconstrained Lagrangian formulation (refer to Basis Pursuit denoising [440]),

$$\min_{\tilde{\boldsymbol{x}}} \|\tilde{\boldsymbol{x}}\|_1 + \lambda \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\Psi}^{-1}\tilde{\boldsymbol{x}}\|_2^2, \tag{18}$$

where $\tilde{\boldsymbol{x}}$ and $\boldsymbol{y}$ denote the transform coefficients and the CS measurements, $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ represent the measurement matrix and the sparsity transform basis respectively, and λ is the Lagrangian multiplier balancing the l_1 norm-driven sparsity over the l_2 norm-based measurement of data fidelity. Many different strategies have been invented for tackling such a convex optimization problem [441], including convex-programming algorithms, gradient descent approaches, greedy methods, and iterative thresholding techniques. Compared to convex-programming-based reconstruction, the other three approaches to CS reconstruction provide remarkable reduction of computational complexity.

5. CONCLUSIONS

In this paper, we present an overview on the development and refinement of both continuous and discrete approaches to energy minimization. These techniques are popularly used in medical applications, such as image segmentation, registration, reconstruction, motion tracking, and compressed sensing. Although the systematic investigation into optimization schemes is growing very quickly, there are still numerous challengeable problems to be tackled. According to the above discussions, we make a few observations listed as follows.

1. Energy minimization has been widely applied in medical image analysis, providing opportunities to handle medical images with various sorts of noises and sampling artifacts in an easier and more meaningful way.
2. The energy function plays a crucial role in energy minimization, which can be formulated from different aspects and minimized using appropriate optimization algorithms such as gradient descent, graph cuts, and primal-dual.

DOI: 10.1002/cnm

3. The energy minimization techniques can be classified into continuous and discrete methods in terms of the space that the energy functionals are optimized over. With respect to different criteria, these optimization approaches can also be categorized in other styles (e.g., local and global, linear and nonlinear, convex and nonconvex), taking into account the objective functionals and/or the collection of constraints.
4. Gradient-based algorithms such as gradient descent, conjugate gradient, proximal gradient are the most prominent continuous optimization techniques in medical imaging, while graph cuts, belief propagation, and linear programming dominate the discrete optimization approaches.
5. Newton's method is computationally expensive due to the evaluation of both first and second derivatives (namely gradients and curvatures) at every iteration. Fortunately, it needs the fewest iterations to find the optimal solution (i.e., rapid convergence). On the contrary, gradient descent methods are considerably faster in every iteration without requiring the computation of second derivatives, but it usually involves many more iterations to reach the minimum (i.e., slow convergence).
6. Conjugate gradient method needs much less computation than Newton's method per iteration without calculating the second derivatives, and it achieves quicker convergence than gradient descent as it involves the previous orientation in the subsequent search.
7. Proximal gradient techniques are popularly applied for tackling convex optimization problems, however their convergence rates still need to be further improved, nevertheless multiple acceleration algorithms have been developed for this purpose.
8. In contrast to many continuous optimization techniques like gradient descent that are prone to local minima, the discrete optimization approaches such as combinatorial graph cuts attempt to obtain global optima. The expansion and swap algorithms are two prevalent graph-cut methods. The former is fairly efficient in practice, but it generally applies to MRFs with metric potential functions, while the latter can handle more generic energy functions as it is applicable to nonmetric potentials as well, nevertheless it is less effective than the expansion algorithm.
9. Tree-reweighted message passing (TRW) together with belief propagation (BP) are two traditional message passing approaches. In comparison with BP, TRW can be utilized to much more energy functions, but its convergence is not ensured.
10. Linear programming (LP) has motivated countless essential theories in optimization like decomposition, convexity as well as duality. The LP problems can be tackled by basis-exchange algorithms (e.g., simplex, criss-cross, conic sampling) together with interior-point methods (e.g., ellipsoid, projective, path-following). However, the former may have worst-case exponential time-complexity, while the latter can achieve a global solution in polynomial time.
11. Primal-dual strategy takes advantage of the duality of an optimization problem, exploiting the information contained in primal variable and dual variable. The dual problem can be tackled by numerous techniques (e.g., gradient descent, interior-point, and Newton's method), which yields a (lower) bound for the primal solution. A variety of primal-dual algorithms have been proposed for solving both continuous and discrete optimization problems with very promising results. However, there are still many problems (e.g., time-complexity, higher-order duality, duality gap) to be addressed.

DOI: 10.1002/cnm

12. Segmenting anatomical structures in the medical images continues to exist as a tough mission because of the large variation in object shapes and variability of image quality. In order to tackle the problems, energy minimization techniques have been substantially investigated and popularly applied in segmenting medical data with encouraging results.
13. Deformable models are one of the earliest developed and most recognized examples of energy minimization techniques being applied to medical image segmentation. The description of object boundaries and applied exterior forces in these models are usually non-identical. The segmentation accuracy and efficiency may vary on different data sets. It is still desirable to design efficient algorithms that are capable of extracting anatomical structures with higher accuracy and more robustness. Fortunately, these approaches are not mutually exclusive, which may be combined together for handling tasks in practice.
14. Gradient-based deformable models generally require careful initial contour placement and computation of derivatives that is susceptible to image noise. They could 'leak' via the object boundaries in case of insufficient remarkable edge features in some specific image regions.
15. Region-based models depend upon the homogeneous properties of spatially localized features. In general, they are more attractive than gradient-based techniques for segmentation, as they are more robust to noise by avoiding derivatives of the image intensity and contour initialization by being less local than most gradient-based approaches.
16. Most gradient-based approaches primarily rely on local information around the active contours, while the region-based methods incline to utilize local information together with global information in the image .
17. For the segmentation of an anatomical structure in a medical image, the prior information of the target shape belonging to this structure could impressively benefit the segmentation procedure. The integration of such priors into the segmentation framework reduces the medical image segmentation complexity.
18. The hybrid deformable models take advantage of the complementary information of gradient-based, region-based, and shape-based methods, thus are more robust to noise and poor boundary initialization.
19. Energy minimization techniques have been vigorously investigated and popularly adopted in medical applications like image segmentation, registration, reconstruction, and motion tracking. The segmentation as well as registration of 2D/3D images allow the construction of anatomical databank, which can possibly be used for exploring the morphological variations, pathological evolution, or growth of organs.
20. By making good use of the complementary information in different imagery, the successful registration of medical images of multi-modality indicates a great potential for surgical implementations such as tumor detection.
21. Compressed sensing is a very promising technique for rapid data acquisition and reconstruction, where energy minimization is involved for the accurate reconstruction of the image from rather fewer measurements in the transform domain. It offers potentially significant reduction of the scan time during medical imaging procedures, which will benefit both patients and healthcare economics.
22. The integration of energy minimization techniques with current medical systems (e.g., disease diagnosis, surgical planning) can additionally validate their implementations in clinical

DOI: 10.1002/cnm

practice and inevitably benefit the advancement of more effective energy minimization schemes that is highly desirable for clinical applications.

23. To carry out a marvelous comparison of various energy minimization methods, it is very necessary to collect some medical data sets as a benchmark and define appropriate protocols for performance evaluation.

REFERENCES

1. Kass M, Witkin A, Terzopoulos D. Snakes: Active contour models. *Int. J. Comput. Vis.* 1988; **1**(4):321–331.
2. McInerney T, Terzopoulos D. Deformable models in medical image analysis. *Proc. Workshop Mathematical Methods in Biomedical Image Analysis*, 1996; 171–180.
3. McInerney T, Terzopoulos D. Deformable models in medical image analysis: A survey. *Med. Image Anal.* 1996; **1**(2):91–108.
4. Xu C, Pham DL, Prince JL. Image segmentation using deformable models. *in: Handbook of Medical Imaging, vol. 2 Medical Image Processing and Analysis (Ch. 3), J. M. Fitzpatrick, M. Sonka (Eds.), SPIE Publications* Jun 2000; :129–174.
5. McIntosh C, Hamarneh G. Medical image segmentation: Energy minimization and deformable models. *in: Medical Imaging: Technology and Applications (Ch. 23), T. Farncombe, K. Iniewski (Eds.), CRC Press* Oct 2013; :661–692.
6. Zhao F, Xie X. An overview on interactive medical image segmentation. *Annals of the BMVA* 2013; **2013**(7):1–22.
7. Cohen LD. On active contour models and balloons. *Computer Vision Graphics and Image Processing: Image Understanding* 1991; **53**(2):211–218.
8. Caselles V, Kimmel R, Sapiro G. Geodesic active contours. *Int. J. Comput. Vis.* 1997; **22**(1):61–79.
9. Bresson X, Esedoglu S, Vandergheynst P, Thiran JP, Osher SJ. Fast global minimization of the active contour/snake model. *J. Math. Imaging Vis.* 2007; **28**(2):151–167.
10. Chan T, Golub G, Mulet P. A nonlinear primal-dual method for total variation-based image restoration. *SIAM J. Sci. Comput.* 1999; **20**(6):1964–1977.
11. Aujol JF, Gilboa G, Chan T, Osher S. Structure-texture image decomposition - modeling, algorithms, and parameter selection. *Int. J. Comput. Vis.* 2006; **67**(1):111–136.
12. Boykov Y, Kolmogorov V. Computing geodesics and minimal surfaces via graph cuts. *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 1, 2003; 26–33.
13. Boykov Y, Funka-Lea G. Graph cuts and efficient N-D image segmentation. *Int. J. Comput. Vis.* 2006; **70**(2):109–131.
14. Szeliski R, Zabih R, Scharstein D, Veksler O, Kolmogorov V, Agarwala A, Tappen M, Rother C. A comparative study of energy minimization methods for Markov random fields. *Proc. European Conf. Comput. Vis.*, 2006; 16–29.
15. Szeliski R, Zabih R, Scharstein D, Veksler O, Kolmogorov V, Agarwala A, Tappen M, Rother C. A comparative study of energy minimization methods for Markov random fields with smoothness-based priors. *IEEE Trans. Pattern Anal. Mach. Intell.* Jun 2008; **30**(6):1068–1080.
16. Elsgolc L. *Calculus of Variations*. Pergamon Press Ltd., London, UK, 1962.
17. Zhu M, Chan TF. An efficient primal-dual hybrid gradient algorithm for total variation image restoration. *Technical Report 08-34*, UCLA CAM 2008.
18. Zhu M, Wright SJ, Chan TF. Duality-based algorithms for total-variation-regularized image restoration. *Computational Optimization and Applications* Nov 2010; **47**(3):377–400.
19. Shi J, Malik J. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* Aug 2000; **22**(8):888–905.
20. Berndt EK, Hall BH, Hall RE, Hausman JA. Estimation and inference in nonlinear structural models. *Annals of Economic and Social Measurement* 1974; **3**(4):653–665.
21. Byrd RH, Lu P, Nocedal J, Zhu C. A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comput.* 1995; **16**(5):1190–1208.
22. Byrd RH, Khalfan HF, Schnabel RB. Analysis of a symmetric rank-one trust region method. *SIAM J. Optimization* 1996; **6**(4):1025–1039.
23. Dence T. Cubics, chaos and Newton's method. *Mathematical Gazette* Nov 1997; **81**(492):403–408.

DOI: 10.1002/cnm

24. Süli E, Mayers D. *An Introduction to Numerical Analysis*. Cambridge University Press, 2003.
25. Nocedal J, Wright SJ. *Numerical Optimization*. 2nd edn., Springer, 2006.
26. Tsai A, Yezzi AJ, Wells W, Tempany C, Tucker D, Fan A, Grimson WE, Willsky A. Model-based curve evolution technique for image segmentation. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 1, 2001; 463–468.
27. Tsai A, Yezzi AJ, Wells W, Tempany C, Tucker D, Fan A, Grimson WE, Willsky A. A shape-based approach to the segmentation of medical imagery using level sets. *IEEE Trans. Med. Imag.* Feb 2003; **22**(2):137–154.
28. Sundaramoorthi G, Yezzi A, Mennucci AC. Sobolev active contours. *Int. J. Comput. Vis.* 2007; **73**(3):345–366.
29. Bar L, Sapiro G. Generalized Newton-type methods for energy formulations in image processing. *SIAM J. Imag. Sci.* 2009; **2**(2):508–531.
30. Hestenes MR, Stiefel E. Methods of conjugate gradients for solving linear systems. *J. Research of the National Bureau of Standards* Dec 1952; **49**(6):409–436.
31. Avriel M. *Nonlinear Programming: Analysis and Methods (Ch. 10 Second derivative, steepest descent, and conjugate gradietn methods)*. Dover Publications, 2003.
32. Knyazev AV, Lashuk I. Steepest descent and conjugate gradient methods with variable preconditioning. *SIAM J. Matrix Analysis and Applications* Nov 2007; **29**(4):1267–1280.
33. Combettes PL, Wajs VR. Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.* 2005; **4**(4):1168–1200.
34. Bioucas-Dias JM, Figueiredo MAT. A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Trans. Image Process.* Dec 2007; **16**(12):2992–3004.
35. Elad M, Matalon B, Zibulevsky M. Coordinate and subspace optimization methods for linear least squares with non-quadratic regularization. *Applied Comput. Harmon. Anal.* Nov 2007; **23**(3):346–367.
36. Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sciences* 2009; **2**(1):183–202.
37. Combettes PL, Pesquet JC. Proximal splitting methods in signal processing. *in: Fixed-Point Algorithms for Inverse Problems in Science and Engineering, Springer Optimization and Its Applications 49 (Ch. 10), H. H. Bauschke et al. (Eds.), Springer* 2011; :185–212.
38. Bezdek JC, Hathaway RJ, Howard RE, Wilson CA, Windham MP. Local convergence analysis of a grouped variable version of coordinate descent. *J. Optimization theory and applications* Sep 1987; **54**(3):471–477.
39. Luo Z, Tseng P. On the convergence of the coordinate descent method for convex differentiable minimization. *J. Optimization theory and applications* Jan 1992; **72**(1):7–35.
40. Tsai A, Wells W, Tempany C, Grimson E, Willsky A. Mutual information in coupled multi-shape model for medical image segmentation. *Med. Image Anal.* 2004; **8**(4):429–445.
41. Elad M. Why simple shrinkage is still relevant for redundant representations? *IEEE Trans. Infor. Theory* Dec 2006; **52**(12):5559–5569.
42. Richtárik P, Takáč M. Parallel coordinate descent methods for big data optimization. *arXiv:1212.0873v2* Nov 2013; :1–43.
43. Hill A, Taylor C. Model-based image interpretation using genetic algorithms. *Image Vision Comput.* 1992; **10**(5):295–300.
44. Tohka J. Global optimization of deformable surface meshes based on genetic algorithms. *Proc. Int. Conf. Image Analysis and Processing*, 2001; 459–464.
45. Fan Y, Jiang T, Evans D. Volumetric segmentation of brain images using parallel genetic algorithms. *IEEE Trans. Med. Imag.* 2002; **21**(8):904–909.
46. McIntosh C, Hamarneh G. Medial-based deformable models in nonconvex shape-spaces for medical image segmentation. *IEEE Trans. Med. Imag.* Jan 2012; **31**(1):33–50.
47. Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* Nov 2001; **23**(11):1222–1239.
48. Kolmogorov V, Zabin R. What energy functions can be minimized via graph cuts? *IEEE Trans. Pattern Anal. Mach. Intell.* 2004; **26**(2):147–159.
49. Grady L. Random walks for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 2006; **28**(11):1768–1783.
50. Komodakis N, Tziritas G. Approximate labeling via graph cuts based on linear programming. *IEEE Trans. Pattern Anal. Mach. Intell.* Aug 2007; **29**(8):1436–1453.
51. Sinop A, Grady L. A seeded image segmentation framework unifying graph cuts and random walker which yields a new algorithm. *Proc. IEEE Int. Conf. Comput. Vis.*, 2007; 1–8.
52. Veksler O. Graph cut based optimization for mrfs with truncated convex priors. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2007; 1–8.

DOI: 10.1002/cnm

53. Pearl J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc., 1988.
54. Yedidia JS, Freeman WT, Weiss Y. Generalized belief propagation. *Advances in Neural Information Processing Systems*, 2001; 689–695.
55. Meltzer T, Yanover C, Weiss Y. Globally optimal solutions for energy minimization in stereo vision using reweighted belief propagation. *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 1, 2005; 428–435.
56. Yedidia JS, Freeman WT, Weiss Y. Constructing free-energy approximations and generalized belief propagation algorithms. *IEEE Trans. Infor. Theory* Jul 2005; **51**(7):2282–2312.
57. Felzenszwalb PF, Huttenlocher DP. Efficient belief propagation for early vision. *Int. J. Comput. Vis.* 2006; **70**(1):41–54.
58. Wainwright MJ, Jaakkola TS, Willsky AS. Map estimation via agreement on trees: Message-passing and linear programming. *IEEE Trans. Infor. Theory* Nov 2005; **51**(11):3697–3717.
59. Kolmogorov V. Convergent tree-reweighted message passing for energy minimization. *IEEE Trans. Pattern Anal. Mach. Intell.* Oct 2006; **28**(10):1568–1583.
60. Dantzig GB. Maximization of a linear function of variables subject to linear inequalities (1947). *in: Activity Analysis of Production and Allocation (Ch. XXI), T. C. Koopmans (Eds.), John Wiley & Sons, Inc. New York, Chapman & Hall Ltd. London* 1951; :339–347.
61. Murty KG. *Linear Programming*. John Wiley & Sons, Inc. New York, 1983.
62. Fukuda K, Terlaky T. Criss-cross methods: A fresh view on pivot algorithms. *Mathematical Programming, Series B* Oct 1997; **79**(1-3):369–395.
63. Chekuri C, Khanna S, Naor J, Zosin L. Approximation algorithms for the metric labeling problem via a new linear programming formulation. *Proc. Annual ACM-SIAM Symp. Discrete Algorithms*, 2001; 109–118.
64. Kleinberg J, Tardos E. Approximation algorithms for classification problems with pairwise relationships: Metric labeling and Markov random fields. *J. ACM* Sep 2002; **49**(5):616–639.
65. Archer A, Fakcharoenphol J, Harrelson C, Krauthgamer R, Talvar K, Tardos E. Approximate classification via earthmover metrics. *Proc. Annual ACM-SIAM Symp. Discrete Algorithms*, 2004; 1079–1087.
66. Serang O. Conic sampling: An efficient method for solving linear and quadratic programming by randomly linking constraints within the interior. *PLOS ONE* Aug 2012; **7**(8):1–12.
67. Taskar B, Guestrin C, Koller D. Max-margin Markov networks. *Advances in Neural Information Processing Systems*, 2004; 25–32.
68. Anguelov D, Taskarf B, Chatalbashev V, Koller D, Gupta D, Heitz G, Ng A. Discriminative learning of Markov random fields for segmentation of 3D scan data. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 2, 2005; 169–176.
69. Taskar B, Chatalbashev V, Koller D, Guestrin C. Learning structured prediction models: A large margin approach. *Proc. Int. Conf. Machine Learning*, 2005; 896–903.
70. Kolmogorov V, Boykov Y, Rother C. Applications of parametric maxflow in computer vision. *Proc. IEEE Int. Conf. Comput. Vis.*, 2007; 1–8.
71. Szummer M, Kohli P, Hoiem D. Learning CRFs using graph cuts. *Proc. European Conf. Comput. Vis.*, 2008; 582–595.
72. Kirkpatrick S, Jr CDG, Vecchi MP. Optimization by simulated annealing. *Science* May 1983; **220**(4598):671–680.
73. Geman S, Geman D. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* Nov 1984; **6**(6):721–741.
74. Barnard ST. Stochastic stereo matching over scale. *Int. J. Comput. Vis.* 1989; **3**(1):17–32.
75. Besag J. On the statistical analysis of dirty pictures. *J. Royal Statistical Soc., Series B* 1986; **48**(3):259–302.
76. Chopp DL. Computing minimal surfaces via level set curvature flow. *J. Comp. Phys.* 1993; **106**(1):77–91.
77. Caselles V, Kimmel R, Sapiro G, Sbert C. Minimal surfaces: A geometric three dimensional segmentation approach. *Numer. Math.* Oct 1997; **77**(4):423–451.
78. Caselles V, Kimmel R, Sapiro G, Sbert C. Minimal surfaces based object segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* Apr 1997; **19**(4):394–398.
79. Appleton B, Talbot H. Globally minimal surfaces by continuous maximal flows. *IEEE Trans. Pattern Anal. Mach. Intell.* 2006; **28**(1):106–118.
80. Grady L. Computing exact discrete minimal surfaces: Extending and solving the shortest path problem in 3D with application to segmentation. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 1, 2006; 69–77.
81. Grady L. Minimal surfaces extend shortest path segmentation methods to 3D. *IEEE Trans. Pattern Anal. Mach. Intell.* Feb 2010; **32**(2):321–334.
82. Mumford D, Shah J. Optimal approximation by piecewise smooth functions and associated variational problems. *Comm. Pure Applied Mathematics* 1989; **42**:577–685.

DOI: 10.1002/cnm

83. Rousson M, Paragios N, Deriche R. Implicit active shape models for 3D segmentation in MRI imaging. *Proc. Med. Image Comput. Comput. Assist. Interv.*, 2004; 209–216.
84. Cremers D. Dynamical statistical shape priors for level set based tracking. *IEEE Trans. Pattern Anal. Mach. Intell.* Aug 2006; **28**(8):1262–1273.
85. Rousson M, Paragios N. Prior knowledge, level set representations & visual grouping. *Int. J. Comput. Vis.* 2008; **76**(3):231–243.
86. Khalfan HF, Byrd RH, Schnabel RB. A theoretical and experimental study of the symmetric rank-one update. *SIAM J. Optimization* 1993; **3**(1):1–24.
87. Bonnans JF, Gilbert JC, Lemaréchal C, Sagastizábal CA. *Numerical Optimization: Theoretical and Practical Aspects*. 2nd edn., Springer, 2006.
88. Nocedal J. Updating Quasi-Newton matrices with limited storage. *Mathematics of Computation* 1980; **35**(151):773–782.
89. Chan TF, Vese LA. Active contours without edges. *IEEE Trans. Image Process.* Feb 2001; **10**(2):266–277.
90. Wright SJ, Nowak RD, Figueiredo MAT. Sparse reconstruction by separable approximation. *IEEE Trans. Signal Process.* Jul 2009; **57**(7):2479–2493.
91. Bauschke HH, Combettes PL. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces (Ch. 19 & 25)*. Springer, 2011.
92. Cremers D, Schmidt F, Barthel F. Shape priors in variational image segmentation: Convexity, lipschitz continuity and globally optimal solutions. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2008; 1–6.
93. Nesterov Y. *Introductory Lectures on Convex Optimization: A Basic Course*. Kluwer Academic Publishers, 2004.
94. Ishikawa H. Higher-order gradient descent by fusion-move graph cut. *Proc. IEEE Int. Conf. Comput. Vis.*, 2009; 568–574.
95. Lempitsky V, Rother C, Blake A. Logcut - efficient graph cut optimization for Markov random fields. *Proc. IEEE Int. Conf. Comput. Vis.*, 2007; 1–8.
96. Golub GH, Overton ML. Convergence of a two-stage Richardson iterative procedure for solving systems of linear equations. *Numerical Analysis, Lecture Notes in Mathematics (Vol. 912)* 1982; :125–139.
97. Golub GH, Ye Q. Inexact preconditioned conjugate gradient method with inner-outer iteration. *SIAM J. Sci. Comput.* 1999; **21**(4):1305–1320.
98. Axelsson O. *Iterative Solution Methods (Ch. 12)*. Cambridge University Press, 1996.
99. Notay Y. Flexible conjugate gradient. *SIAM J. Sci. Comput.* 2000; **22**(4):1444–1460.
100. Shewchuk JR. An introduction to the conjugate gradient method without the agonizing pain. *Technical Report CMUCS-TR-94-125*, Carnegie Mellon University 1994.
101. Chambolle A, DeVore RA, Lee NY, Lucier BJ. Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage. *IEEE Trans. Image Process.* Mar 1998; **7**(3):319–335.
102. Daubechies I, Defrise M, Mol CD. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Applied Math.* Nov 2004; **57**(11):1413–1457.
103. Vonesch C, Unser M. A fast thresholded Landweber algorithm for wavelet-regularized multidimensional deconvolution. *IEEE Trans. Image Process.* Apr 2008; **17**(4):539–549.
104. Landweber L. An iteration formula for Fredholm integral equations of the first kind. *American J. Math.* Jul 1951; **73**(3):615–624.
105. Hanke M, Neubauer A, Scherzer O. A convergence analysis of the Landweber iteration for nonlinear ill-posed problems. *Numer. Math.* Nov 1995; **72**(1):21–37.
106. von Neumann J. On rings of operators. reduction theory. *Annals of Math.* Apr 1949; **50**(2):401–485.
107. Deutsch F. *Best Approximation in Inner Product Spaces (Ch. 9)*. Springer, 2001.
108. Glowinski R, Le Tallec P. *Augmented Lagrangian and operator splitting methods in nonlinear mechanics*. SIAM Studies in Applied Mathematics, 1989.
109. Eckstein J, Bertsekas DP. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming* Apr 1992; **55**(1-3):293–318.
110. Goldstein T, Osher S. The split Bregman method for l1 regularized problems. *SIAM J. Imaging Sci.* Apr 2009; **2**(2):323–343.
111. Björck A. *Numerical Methods for Least Squares Problems (Ch. 7)*. SIAM, 1996.
112. Johansson B, Elfving T, Kozlovc V, Censor Y, Forssen PE, Granlund G. The application of an oblique-projected Landweber method to a model of supervised learning. *Math. Comput. Modelling* Apr 2006; **43**(7-8):892–909.
113. Bauschke HH, Borwein JM. On the convergence of von Neumann's alternating projection algorithm for two sets. *Set-Valued Analysis* 1993; **1**(2):185–212.

DOI: 10.1002/cnm

114. Boyle JP, Dykstra RL. A method for finding projections onto the intersection of convex sets in Hilbert spaces. *Advances in Order Restricted Statistical Inference, R. L. Dykstra et al. (Eds.), Lecture Notes in Statistics 37, Springer Verlag*, 1986; 28–47.
115. Gaffke N, Mathar R. A cyclic projection algorithm via duality. *Metrika* 1989; **36**(1):29–54.
116. Lewis AS, Malick J. Alternating projections on manifolds. *Mathematics of Operations Research* Feb 2008; **33**(1):216–234.
117. Bauschke HH, Borwein JM. On projection algorithms for solving convex feasibility problems. *SIAM Review* Sep 1996; **38**(3):367–426.
118. Escalante R, Raydan M. *Alternating Projection Methods*. SIAM, 2011.
119. Trussell HJ, Civanlar MR. The Landweber iteration and projection onto convex sets. *IEEE Trans. Acoust. Speech Signal Process.* Dec 1985; **33**(6):1632–1634.
120. Fukushima M. Application of the alternating direction method of multipliers to separable convex programming problems. *Computational Optimization and Applications* Oct 1992; **1**(1):93–11.
121. Boyd S, Parikh N, Chu E, Peleato B, Eckstein J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning* Jan 2011; **3**(1):1–122.
122. Tosserams S, Etman LFP, Papalambros PY, Rooda JE. An augmented Lagrangian relaxation for analytical target cascading using the alternating direction method of multipliers. *Structural and Multidisciplinary Optimization* Mar 2006; **31**(3):176–189.
123. Hong M, Luo ZQ. On the linear convergence of the alternating direction method of multipliers. *arXiv:1208.3922v3* Mar 2013; :1–37.
124. Bregman LM. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics* 1967; **7**(3):200–217.
125. Goldstein T, Bresson X, Osher S. Geometric applications of the split Bregman method: Segmentation and surface reconstruction. *J. Scientific Computing* Oct 2010; **45**(1-3):272–293.
126. Paul G, Cardinale J, Sbalzarini IF. An alternating split Bregman algorithm for multi-region segmentation. *Asilomar Proc. Signals, Systems and Computers*, 2011; 426–430.
127. Yin W. Analysis and generalizations of the linearized Bregman method. *SIAM J. Imaging Sciences* Oct 2010; **3**(4):856–877.
128. Yang Y, Möller M, Osher S. A dual split Bregman method for fast l1 minimization. *Math. Computation* May 2013; **82**(284):2061–2085.
129. Huang B, Ma S, Goldfarb D. Accelerated linearized bregman method. *J. Scientific Computing* Feb 2013; **54**(2-3):428–453.
130. Moradifam A, Nachman A. Convergence of the alternating split Bregman algorithm in infinite-dimensional Hilbert spaces. *arXiv:1112.1960v1* Dec 2011; :1–14.
131. Esser E. Applications of Lagrangian-based alternating direction methods and connections to split Bregman. *Technical Report 09-31*, University of California 2009.
132. Mosci S, Rosasco L, Santoro M, Verri A, Villa S. Solving structured sparsity regularization with proximal methods. *Machine Learning and Knowledge Discovery in Databases, J. L. Balcázar et al. (Eds.), LNCS 6322, Springer Verlag*, 2010; 418–433.
133. Loshchilov I, Schoenauer M, Sebag M. Adaptive coordinate descent. *Proc. Annual Conf. Genetic and evolutionary computation*, 2011; 885–892.
134. Tseng P. Convergence of block coordinate descent method for nondifferentiable minimization. *J. Optimization Theory and Applications* Jun 2001; **109**(3):475–494.
135. Hansen N. Adaptive encoding: How to render search coordinate system invariant. *Parallel Problem Solving from Nature, G. Rudolph et al. (Eds.), LNCS 5199, Springer Verlag*, 2008; 205–214.
136. Hansen N, Ostermeier A. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation* 2001; **9**(2):159–195.
137. Igel C, Hansen N, Roth S. Covariance matrix adaptation for multi-objective optimization. *Evolutionary Computation* 2007; **15**(1):1–28.
138. Loshchilov I, Schoenauer M, Sebag M. Comparison-based optimizers need comparison-based surrogates. *Parallel Problem Solving from Nature, R. Schaefer et al. (Eds.), LNCS 6238, Springer Verlag*, 2010; 364–373.
139. Ibánez O, Barreira N, Santos J, Penedo M. Genetic approaches for topological active nets optimization. *Pattern Recognit.* 2009; **42**(5):907–917.
140. Ballerini L. Genetic snakes for medical image segmentation. *Proc. European Workshops Evolutionary Image Analysis, Signal Processing and Telecommunications*, vol. 1596, 1999; 59–73.

DOI: 10.1002/cnm

141. Williams DJ, Shah M. A fast algorithm for active contours and curvature estimation. *CVGIP: Imag. Under.* 1992; **55**(1):14–26.
142. Dantzig GB, Thapa MN. *Linear programming 1: Introduction*. Springer-Verlag, 1997.
143. Dantzig GB, Thapa MN. *Linear programming 2: Theory and Extensions*. Springer-Verlag, 2003.
144. Nikolova M, Esedoglu S, Chan TF. Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM J. Applied Mathematics* 2006; **66**(5):1632–1648.
145. Andrews S, McIntosh C, Hamarneh G. Convex multi-region probabilistic segmentation with shape prior in the isometric log-ratio transformation space. *Proc. IEEE Int. Conf. Comput. Vis.*, 2011; 2096–2103.
146. Barrett WA, Mortensen EN. Interactive live-wire boundary extraction. *Medical Image Analysis* 1997; **1**(4):331–341.
147. Wink O, Niessen WJ, Viergever M. Multiscale vessel tracking. *IEEE Trans. Med. Imag.* 2004; **23**(1):130–133.
148. Poon M, Hamarneh G, Abugharbieh R. Efficient interactive 3d livewire segmentation of objects with arbitrarily topologies. *Computerized Medical Imaging and Graphics* 2008; **32**(8):639–650.
149. Kawahara J, McIntosh C, Tam R, Hamarneh G. Globally optimal spinal cord segmentation using a minimal path in high dimensions. *Proc. IEEE Int. Symp. Biomedical Imaging: From Nano to Macro*, 2013; 848–851.
150. Dijkstra EW. A note on two problems in connexion with graphs. *Numer. Math.* 1959; **1**(1):269–271.
151. Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts. *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 1, 1999; 377–384.
152. Boykov Y, Jolly MP. Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. *Proc. IEEE Int. Conf. Comput. Vis.*, 2001; 105–112.
153. Komodakis N, Tziritas G. A new framework for approximate labeling via graph cuts. *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2, 2005; 1018–1025.
154. Sun J, Shum HY, Zheng NN. Stereo matching using belief propagation. *Proc. European Conf. Comput. Vis.*, 2002; 510–524.
155. Sun J, Zheng NN, Shum HY. Stereo matching using belief propagation. *IEEE Trans. Pattern Anal. Mach. Intell.* Jul 2003; **25**(7):787–800.
156. Sarkis M, Diepold K. Sparse stereo matching using belief propagation. *Proc. IEEE Int. Conf. Image Processing*, 2008; 1780–1783.
157. Wainwright MJ, Jaakkola TS, Willsky AS. Exact MAP estimates by (hyper) tree agreement. *Advances in Neural Information Processing Systems*, 2003; 809–816.
158. Lovász L. On the ratio of optimal integral and fractional covers. *Discrete Mathematics* 1975; **13**(4):383–390.
159. Felzenszwalb PF, Huttenlocher DP. Efficient belief propagation for early vision. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 1, 2004; 261–268.
160. Dantzig GB. *Linear Programming and Extensions (Ch. 5)*. Princeton University Press, 1963.
161. Terlaky T, Zhang S. Pivot rules for linear programming: A survey on recent theoretical developments. *Annals of Operations Research* 1993; **46-47**(1):203–233.
162. Khachiyan LG. A polynomial algorithm for linear programming. *Soviet Math. Doklady* 1979; **20**:191–194.
163. Karmarkar N. A new polynomial-time algorithm for linear programming. *Combinatorica* 1984; **4**:373–395.
164. Gondzio J, Terlaky T. A computational view of interior point methods. *in: Advances in linear and integer programming (Ch. 3), J. E. Beasley (Eds.), Oxford University Press* 1996; :103–144.
165. Padberg MW. *Linear Optimization and Extensions*. 2nd edn., Springer-Verlag, 1999.
166. Bland RG. New finite pivoting rules for the simplex method. *Math. Oper. Res.* May 1977; **2**(2):103–107.
167. Klee V, Minty GJ. How good is the simplex algorithm? *Proc. Symp. Inequalities*, 1972; 159–175.
168. Schrijver A. *Theory of Linear and Integer Programming*. John Wiley & Sons, 1998.
169. Maros I, Mitra G. Simplex algorithms. *in: Advances in linear and integer programming (Ch. 1), J. E. Beasley (Eds.), Oxford University Press* 1996; :1–46.
170. Alevras D, Padberg MW. *Linear Optimization and Extensions: Problems and Solutions*. Springer-Verlag, 2001.
171. Vanderbei RJ. *Linear Programming: Foundations and Extensions*. 4th edn., Springer, 2013.
172. Beasley JE. *Advances in Linear and Integer Programming*. Oxford University Press, 1996.
173. Wright SJ. *Primal-Dual Interior-Point Methods*. SIAM, 1997.
174. Barnes ER. A variation on karmarkar's algorithm for solving linear programming problems. *Mathematical Programming* Nov 1986; **36**(2):174–182.
175. Bajalinov EB. *Linear-Fractional Programming: Theory, Methods, Applications and software*. Kluwer Academic Publishers, 2003.
176. Craven BD. *Fractional programming (Ch. 5)*. Heldermann Verlag, Berlin, 1988.
177. Illés T, Szirmai Á, Terlaky T. The finite criss-cross method for hyperbolic programming. *European J. Operational Research* Apr 1999; **114**(1):198–214.

178. Mortensen EN, Barrett WA. Interactive segmentation with intelligent scissors. *Graphical Models and Image Processing* 1998; **60**(5):349–384.
179. Yezzi AJ, Kichenassamy S, Kumar A, Olver P, Tannenbaum A. A geometric snake model for segmentation of medical imagery. *IEEE Trans. Med. Imag.* Apr 1997; **16**(2):199–209.
180. Bates P, Wei GW, Zhao S. Minimal molecular surfaces and their applications. *J. Comput. Chem.* 2008; **29**(3):380–391.
181. Hu L, Chen D, Wei GW. High-order fractional partial differential equation transform for molecular surface construction. *Mol. Based Math. Biol.* Jan 2013; **1**:1–25.
182. Boyd S, Vandenberghe L. *Convex Optimization (Ch. 5)*. Cambridge University Press, 2004.
183. Rockafellar RT. *Convex Analysis (Part VI)*. Princeton University Press, 1970.
184. Wolfe P. A duality theorem for non-linear programming. *Quarterly of Applied Math.* 1961; **19**:239–244.
185. Geoffrion AM. Duality in nonlinear programming: A simplified applications-oriented development. *SIAM Review* 1971; **13**(1):1–37.
186. Mond B, Weir T. Generalized concavity and duality. *in: Generalized Concavity in Optimization and Economics, S. Schaible and W. T. Ziemba (Eds.), Academic Press* 1981; :263–279.
187. Mond B. Mond-Weir duality. *in: Structure and Applications, Springer Optimization and Its Applications, Vol. 32 (Ch. 8), C. Pearce and E. Hunt (Eds.), Springer* 2009; :157–165.
188. Fenchel W. *Convex Cones, Sets and Functions*. Princeton University, 1951.
189. Rockafellar RT. Extension of Fenchel's duality theorem for convex functions. *Duke Math. J.* 1966; **33**(1):1–245.
190. Aubin JP, Ekeland I. Estimates of the duality gap in nonconvex optimization. *Math. Oper. Res.* Aug 1976; **1**(3):225–245.
191. Giannessi F. On the theory of Lagrangian duality. *Optim. Lett.* Jan 2007; **1**(1):9–20.
192. Giannessi F, Mastroeni G. Separation of sets and Wolfe duality. *J. Glob. Optim.* Nov 2008; **42**(3):401–412.
193. Ye X, Han SP, Lin A. A note on the connection between the primal-dual and the A* algorithm. *Int. J. Operations Research and Information Systems* 2010; **1**(1):73–85.
194. Kuhn HW, Tucker AW. Nonlinear programming. *Proc. 2nd Berkeley Symp.*, 1951; 481–492.
195. Borwein JM, Lewis AS. *Convex Analysis and Nonlinear Optimization: Theory and Examples, 2nd ed.* Springer, 2006.
196. Mangasarian OL. Second- and higher-order duality in nonlinear programming. *J. Math. Anal. Appl.* Sep 1975; **51**(3):607–620.
197. Ivanov VI. On the optimality of some classes of invex problems. *Optim. Lett.* Jan 2012; **6**(1):43–54.
198. Ivanov VI. Duality in nonlinear programming. *Optim. Lett.* Dec 2013; **7**(8):1643–1658.
199. Rudin LI, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms. *Phys. D* 1992; **60**(1-4):259–268.
200. Chambolle A. An algorithm for total variation minimization and applications. *J. Math. Imaging Vis.* 2004; **20**(1-2):89–97.
201. Goldfarb D, Yin W. Second-order cone programming methods for total variation-based image restoration. *SIAM J. Sci. Comput.* 2005; **27**(2):622–645.
202. Collette Y, Siarry P. *Multiobjective Optimization: Principles and Case Studies*. Springer, Berlin, Germany, 2002.
203. Hanning T, Schöne R, Pisinger G. Vector image segmentation by piecewise continuous approximation. *J. Math. Imaging Vis.* Jul 2006; **25**(1):5–23.
204. Nakib A, Oulhadj H, Siarry P. Image thresholding based on pareto multiobjective optimization. *Engineering Applications of Artificial Intell.* Apr 2010; **23**(3):313–320.
205. Tappen MF, Freeman WT. Comparison of graph cuts with belief propagation for stereo, using identical MRF parameters. *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 1, 2003; 900–906.
206. Boykov Y, Kolmogorov V. An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision. *IEEE Trans. Pattern Anal. Mach. Intell.* Sep 2004; **26**(9):1124–1137.
207. Kolmogorov V, Rother C. Comparison of energy minimization algorithms for highly connected graphs. *Proc. European Conf. Comput. Vis.*, 2006; 1–15.
208. CRAN. http://cran.r-project.org/view=Optimization.
209. Wikipedia. https://en.wikipedia.org/wiki/List_of_optimization_software.
210. Wikipedia. https://en.wikipedia.org/wiki/Comparison_of_optimization_software.
211. ADMB. http://admb-project.org.
212. CVXOPT. http://cvxopt.org.
213. NLopt. http://ab-initio.mit.edu/wiki/index.php/NLopt.
214. Optimx. http://cran.r-project.org/web/packages/optimx.
215. pyOpt. http://www.pyopt.org.

DOI: 10.1002/cnm

216. LINGO. http://www.lindo.com/products/lingo.
217. Maple. http://www.maplesoft.com/support/help/maple/view.aspx?path=Optimization.
218. MATLAB. http://www.mathworks.com/products/optimization.
219. Mathematica. https://reference.wolfram.com/language/guide/Optimization.html.
220. Singh A, Goldgof D, Terzopoulos D. *Deformable models in medical image analysis*. Los Alamitos, CA: IEEE Computer Society, 1998.
221. Staib LH, Duncan JS. Boundary finding with parametrically deformable models. *IEEE Trans. Pattern Anal. Mach. Intell.* Nov 1992; **14**(11):1061–1075.
222. Chakraborty A, Staib LH, Duncan JS. An integrated approach to boundary finding in medical images. *Proc. IEEE Workshop Biomedical Image Analysis*, 1994; 13–22.
223. Wang Y, Staib LH. Boundary finding with correspondence using statistical shape models. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 1998; 338–345.
224. Leventon ME, Grimson WEL, Faugeras O. Statistical shape influence in geodesic active contours. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 1, 2000; 316–323.
225. Zeng X, Staib LH, Schultz RT, Duncan JS. Segmentation and measurement of the cortex from 3-D MR images using coupled-surfaces propagation. *IEEE Trans. Med. Imag.* Oct 1999; **18**(10):927–937.
226. Li Y, Sun J, Tang CK, Shum HY. Lazy snapping. *Proc. ACM SIGGRAPH*, 2004; 303–308.
227. Rother C, Kolmogorov V, Blake A. Grabcut - Interactive foreground extraction using iterated graph cuts. *Proc. ACM SIGGRAPH*, 2004; 309–314.
228. Armstrong CJ, Price BL, Barrett WA. Interactive segmentation of image volumes with live surface. *Computers & Graphics* 2007; **31**(2):212–229.
229. Essa E, Xie X, Sazonov I, Nithiarasu P. Automatic IVUS media-adventitia border extraction using double interface graph cut segmentation. *Proc. IEEE Int. Conf. Image Processing*, 2011; 69–72.
230. Essa E, Xie X, Sazonov I, Nithiarasu P, Smith D. Shape prior model for media-adventitia border segmentation in IVUS using graph cut. *Proc. MICCAI Medical Computer Vision*, 2012; 98–107.
231. Jones JL, Essa E, Xie X, Smith D. Interactive segmentation of media-adventitia border in IVUS. *Proc. Int. Conf. Computer Analysis of Images and Patterns*, 2013; 466–474.
232. Jones JL, Xie X, Essa E. Combining region-based and imprecise boundary-based cues for interactive medical image segmentation. *Int. J. Numerical Methods in Biomed. Eng.* Dec 2014; **30**(12):1649–1666.
233. Li K, Wu X, Chen DZ, Sonka M. Optimal surface segmentation in volumetric images-A graph-theoretic approach. *IEEE Trans. Pattern Anal. Mach. Intell.* 2006; **28**(1):119–134.
234. Yuan X, Zhang N, Nguyen MX, Chen B. Volume cutout. *Visual Computer (Special Issue of Pacific Graphics)* 2005; **21**(8-10):745–754.
235. Lombaert H, Sun Y, Grady L, Xu C. A multilevel banded graph cuts method for fast image segmentation. *Proc. IEEE Int. Conf. Comput. Vis.*, 2005; 259–265.
236. Essa E, Xie X, Jones JL. Graph based lymphatic vessel wall localisation and tracking. *Proc. 10th IAPR-TC15 Workshop Graph-Based Representations in Pattern Recognition*, 2015; 345–354.
237. Essa E, Xie X, Jones JL. Minimum s-excess graph for segmenting and tracking multiple borders with HMM. *Proc. Int. Conf. Medical Image Computing and Computer Assisted Intervention*, 2015.
238. Jones JL, Essa E, Xie X. Automatic segmentation of lymph vessel wall using optimal surface graph cut and hidden Markov models. *Proc. 37th Annual Int. Conf. IEEE Engineering in Medicine and Biology Society*, 2015.
239. Osher SJ, Sethian JA. Fronts propagation with curvature dependent speed: Algorithms based on Hamilton-Jacobi formulations. *J. Comp. Phys.* 1988; **79**:12–49.
240. Kichenassamy S, Kumar A, Olver P, Tannenbaum A, Yezzi A. Gradient flows and geometric active contour models. *Proc. IEEE Int. Conf. Comput. Vis.*, 1995; 810–815.
241. Sethian JA. *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Material Science*. Cambridge University Press, 1999.
242. Lorigo LM, Faugeras O, Grimson W, Keriven R, Kikinis R, Westin CF. Co-dimension 2 geodesic active contours for MRA segmentation. *Proc. Int. Conf. Inf. Proc. in Med. Imaging*, 1999; 126–139.
243. Yezzi A, Tsai A, Willsky A. A statistical approach to snakes for bimodal and trimodal imagery. *Proc. IEEE Int. Conf. Comput. Vis.*, 1999; 898–903.
244. Guo Y, Vemuri B. Hybrid geometric active models for shape recovery in medical images. *Proc. Int. Conf. Inf. Proc. in Med. Imaging*, 1999; 112–125.
245. Yeo SY, Xie X, Sazonov I, Nithiarasu P. Level set segmentation with robust image gradient energy and statistical shape prior. *Proc. IEEE Int. Conf. Image Processing*, 2011; 3397–3400.
246. Paiement A, Mirmehdi M, Xie X, Hamilton M. Integrated segmentation and interpolation of sparse data. *IEEE Trans. Imag. Proc.* Jan 2014; **23**(1):110–125.

DOI: 10.1002/cnm

247. Yezzi A, Tsai A, Willsky A. A fully global approach to image segmentation via coupled curve evolution equations. *J. Visual Communication and Image Representation* Mar 2002; **13**(1-2):195–216.
248. Xie X, Mirmehdi M. MAC: Magnetostatic active contour model. *IEEE Trans. Pattern Anal. Mach. Intell.* 2008; **30**(4):632–647.
249. Kim J, Fisher JW, Yezzi A, Cetin M, Willsky AS. Nonparametric methods for image segmentation using information theory and curve evolution. *Proc. IEEE Int. Conf. Image Processing*, vol. 3, 2002; 797–800.
250. Kim J, Fisher JW, Yezzi A, Cetin M, Willsky AS. A nonparametric statistical method for image segmentation using information theory and curve evolution. *IEEE Trans. Image Process.* Oct 2005; **14**(10):1486–1502.
251. Xie X, Mirmehdi M. Implicit active model using radial basis function interpolated level sets. *Proc. British Mach. Vis. Conf.*, 2007; 1040–1049.
252. Chan TF, Vese LA. An efficient variational multiphase motion for the Mumford-Shah segmentation model. *Proc. Asilomar Conf. Signals, Systems and Computers*, vol. 1, 2000; 490–494.
253. Sethian JA. *Level Set Methods: Evolving Interfaces in Geometry, Fluid Mechanics, Computer Vision, and Material Science*. Cambridge University Press, 1996.
254. Siddiqi K, Lauzière Y, Tannenbaum A, Zucker S. Area and length minimizing flows for shape segmentation. *IEEE Trans. Image Process.* 1998; **7**(3):433–443.
255. McInerney T, Terzopoulos D. Topology adaptive deformable surfaces for medical image volume segmentation. *IEEE Trans. Med. Imag.* Oct 1999; **18**(10):840–850.
256. Xu C, Prince JL. Generalized gradient vector flow external forces for active contours. *Signal Process.* 1998; **71**(2):131–139.
257. Xu C, Prince JL. Snakes, shapes, and gradient vector flow. *IEEE Trans. Image Process.* Mar 1998; **7**(3):359–369.
258. Cohen LD, Cohen I. Finite-element methods for active contour models and balloons for 2-D and 3-D images. *IEEE Trans. Pattern Anal. Mach. Intell.* 1993; **15**(11):1131–1147.
259. MacDonald D, Avis D, Evans AC. Multiple surface identification and matching in magnetic resonance images. *Proc. SPIE Conf. Visualization in Biomedical Computing*, vol. 2359, 1994; 160–169.
260. Gil D, Radeva P. Curvature vector flow to assure convergent deformable models for shape modelling. *Proc. Int. Workshop Energy Minimization Methods in Computer Vision and Pattern Recognition*, 2003; 357–372.
261. Xie X, Mirmehdi M. RAGS: Region-aided geometric snake. *IEEE Trans. Image Process.* May 2004; **13**(5):640–652.
262. Jalba A, Wilkinson M, Roerdink J. CPM: A deformable model for shape recovery and segmentation based on charged particles. *IEEE Trans. Pattern Anal. Mach. Intell.* Oct 2004; **26**(10):1320–1335.
263. Paragios N, Mellina-Gottardo O, Ramesh V. Gradient vector flow fast geometric active contours. *IEEE Trans. Pattern Anal. Mach. Intell.* 2004; **26**(3):402–407.
264. Yeo SY, Xie X, Sazonov I, Nithiarasu P. Geometric potential force for the deformable model. *Proc. British Mach. Vis. Conf.*, 2009; 1–11.
265. Yeo SY, Xie X, Sazonov I, Nithiarasu P. Geometrically induced force interaction for three-dimensional deformable models. *IEEE Trans. Image Process.* May 2011; **20**(5):1373–1387.
266. Liang J, McInerney T, Terzopoulos D. United snakes. *Med. Image Anal.* 2006; **10**(2):215–233.
267. Paragios N, Deriche R. Geodesic active regions and level set methods for supervised texture segmentation. *Int. J. Comput. Vis.* 2002; **46**(3):223–247.
268. Cremers D, Rousson M, Deriche R. A review of statistical approaches to level set segmentation: Integrating color, texture, motion and shape. *Int. J. Comput. Vis.* 2007; **72**(2):195–215.
269. Cootes TF, Cooper D, Taylor CJ, Graham J. Active shape models - their training and application. *Comput. Vis. Image Understand.* Jan 1995; **61**(1):38–59.
270. Cootes TF, Taylor CJ. A mixture model for representing shape variation. *Imag. Vis. Comp.* Jun 1999; **17**(8):567–573.
271. Cootes TF, Edwards GJ, Taylor CJ. Active appearance models. *IEEE Trans. Pattern Anal. Mach. Intell.* 2001; **23**(6):681–685.
272. Bresson X, Vandergheynst P, Thiran JP. A variational model for object segmentation using boundary information and shape prior driven by the Mumford-Shah functional. *Int. J. Comput. Vis.* 2006; **68**(2):145–162.
273. Malladi R, Sethian JA, Vemuri BC. Shape modeling with front propagation: A level set approach. *IEEE Trans. Pattern Anal. Mach. Intell.* Feb 1995; **17**(2):158–175.
274. Caselles V, Catte F, Coll T, Dibos F. A geometric model for active contours in image processing. *Numer. Math.* 1993; **66**(1):1–31.
275. Xie X. Active contouring based on gradient vector interaction and constrained level set diffusion. *IEEE Trans. Image Process.* 2010; **19**(1):154–164.

DOI: 10.1002/cnm

276. Tsai A, Yezzi AJ, Willsky AS. Curve evolution implementation of the Mumford-Shah functional for image segmentation, denoising, interpolation, and magnification. *IEEE Trans. Image Process.* Aug 2001; **10**(8):1169–1186.
277. Chan TF, Moelich M, Sandberg B. Some recent developments in variational image segmentation. *Image Processing Based on Partial Differential Equations, Springer, Berlin, Germany* 2007; :175–210.
278. Cootes TF, Taylor CJ, Cooper DH, Graham J. Training models of shape from sets of examples. *Proc. British Mach. Vis. Conf.*, 1992; 9–18.
279. Warfield SK, Kaus M, Jolesz FA, Kikinis R. Adaptive, template moderated, spatially varying statistical classification. *Med. Image Anal.* 2000; **4**(1):43–55.
280. Cremers D, Kohlberger T, Schnörr C. Nonlinear shape statistics in Mumford-Shah based segmentation. *Proc. European Conf. Comput. Vis.*, 2002; 93–108.
281. Dambreville S, Rathi Y, Tannen A. Shape-based approach to robust image segmentation using kernel PCA. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 1, 2006; 977–984.
282. Etyngier P, Segonne F, Keriven R. Shape priors using manifold learning techniques. *Proc. IEEE Int. Conf. Comput. Vis.*, 2007; 1–8.
283. Pohl KM, Fisher J, Bouix S, Shenton M, McCarley RW, Grimson WEL, Kikinis R, Wells WM. Using the logarithm of odds to define a vector space on probabilistic atlases. *Med. Image Anal.* 2007; **11**(5):465–477.
284. Duta N, Sonka M. Segmentation and interpretation of mr brain images: an improved active shape model. *IEEE Trans. Med. Imag.* Dec 1998; **17**(6):1049–1062.
285. Frangi AF, Rueckert D, Schnabel JA, Niessen WJ. Automatic construction of multiple-object three-dimensional statistical shape models: application to cardiac modeling. *IEEE Trans. Med. Imag.* Sep 2002; **21**(9):1151–1166.
286. Delingette H, Hebert M, Ikeuchi K. Shape representation and image segmentation using deformable surfaces. *Imag. Vis. Comp.* 1992; **10**(3):132–144.
287. McInerney T, Terzopoulos D. A dynamic finite element surface model for segmentation and tracking in multidimensional medical images with application to cardiac 4D image analysis. *Comp. Med. Imag. Graph.* 1995; **19**(1):69–83.
288. Montagnat J, Delingette H, Ayache N. A review of deformable surfaces: Topology, geometry and deformation. *Imag. Vis. Comp.* 2001; **19**(14):1023–1040.
289. Pizer SM, Fletcher PT, Joshi S, Thall A, Chen JZ, Fridman Y, Fritsch DS, Gash AG, Glotzer JM, Jiroutek MR, *et al.*. Deformable M-Reps for 3D medical image segmentation. *Int. J. Comput. Vis.* 2003; **55**(2-3):85–106.
290. McInerney T, Terzopoulos D. Topologically adaptable snakes. *Proc. IEEE Int. Conf. Comput. Vis.*, 1995; 840–845.
291. Delingette H. General object reconstruction based on simplex meshes. *Int. J. Comput. Vis.* 1999; **32**(2):111–146.
292. McInerney T, Terzopoulos D. T-snakes: Topology adaptive snakes. *Med. Image Anal.* Jun 2000; **4**(2):73–91.
293. Falcão AX, Udupa JK. Segmentation of 3D objects using live wire. *Proc. SPIE Conf. Medical Imaging: Image Processing (Ed. K. M. Hanson)*, vol. 3034, 1997; 228–235.
294. McIntosh C, Hamarneh G. Vessel crawlers: 3D physically-based deformable organisms for vasculature segmentation and analysis. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 1, 2006; 1084–1091.
295. Paragios N, Deriche R. Coupled geodesic active regions for image segmentation: A level set approach. *Proc. European Conf. Comput. Vis.*, 2000; 224–240.
296. Vese LA, Chan T. A multiphase level set framework for image segmentation using the mumford and shah model. *Int. J. Comput. Vis.* 2002; **50**(3):271–293.
297. Grady L, , Funka-Lea G. Multi-label image segmentation for medical applications based on graph-theoretic electrical potentials. *Proc. ECCV Workshop Computer Vision Approaches to Medical Image Analysis and Mathematical Methods in Biomedical Image Analysis*, 2004; 230–245.
298. Chakraborty A, Staib LH, Duncan JS. Deformable boundary finding in medical images by integrating gradient and region information. *IEEE Trans. Med. Imag.* 1996; **15**(6):859–870.
299. Cootes TF, Beeston C, Edwards GJ, Taylor CJ. A unified framework for atlas matching using active appearance models. *Proc. Int. Conf. Inf. Proc. in Med. Imaging*, 1999; 322–333.
300. Kimmel R. Fast edge integration. *in: Geometric Level Set Methods in Imaging, Vision, and Graphics (Ch. 4), S. Osher, N. Paragios (Eds.), Berlin, Germany: Springer-Verlag* Aug 2003; :59–77.
301. McIntosh C, Hamarneh G. Is a single energy functional sufficient? Adaptive energy functionals and automatic initialization. *Proc. Int. Conf. Medical Image Computing and Computer Assisted Intervention*, 2007; 503–510.
302. McIntosh C, Hamarneh G. Optimal weights for convex functionals in medical image segmentation. *Proc. Int. Symp. Visual Computing, Special Track on Optimization for Vision, Graphics and Medical Imaging: Theory and Applications*, 2009; 1079–1088.
303. Rao J, Abugharbieh R, Hamarneh G. Adaptive regularization for image segmentation using local image curvature cues. *Proc. European Conf. Comput. Vis.*, 2010; 651–665.

DOI: 10.1002/cnm

304. Cover T, Thomas J. *Elements of Information Theory, 1st ed.* Wiley, New York, 1991.
305. Eviatar H, Somorjai RL. A fast simple active contour algorithm for biomedical images. *Pattern Recognit. Letters* 1996; **17**:969–974.
306. Delingette H, Montagnat J. Shape and topology constraints on parametric active contours. *Comput. Vis. Image Understand.* 2001; **83**(2):140–171.
307. Amini AA, Weymouth TE, Jain RC. Using dynamic programming for solving variational problems in vision. *IEEE Trans. Pattern Anal. Mach. Intell.* 1990; **12**(9):855–867.
308. Terzopoulos D, Witkin A, Kass M. Constraints on deformable models: recovering 3D shape and nonrigid motion. *Artificial Intell.* 1988; **36**(1):91–123.
309. Terzopoulos D, Metaxas D. Dynamic 3D models with local and global deformations: deformable superquadrics. *IEEE Trans. Pattern Anal. Mach. Intell.* 1991; **13**(7):703–714.
310. Bredno J, Lehmann TM, Spitzer K. A general discrete contour model in two, three, and four dimensions for topology-adaptive multichannel segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* May 2003; **25**(5):550–563.
311. Lauchaud JO, Taton B. Deformable model with a complexity independent from image resolution. *Comput. Vis. Image Understand.* 2005; **99**(3):453–475.
312. Sapiro G, Tannenbaum A. Affine invariant scale-space. *Int. J. Comput. Vis.* 1993; **11**(1):25–44.
313. Kimmel R, Amir A, Bruckstein AM. Finding shortest paths on surfaces using level sets propagation. *IEEE Trans. Pattern Anal. Mach. Intell.* 1995; **17**(6):635–640.
314. Baillard C, Barillot C. Robust 3D segmentation of anatomical structures with level sets. *Proc. Int. Conf. Medical Image Computing and Computer Assisted Intervention*, 2000; 236–245.
315. Holtzman-Gazit M, Kimmel R, Peled N, Goldsher D. Segmentation of thin structures in volumetric medical images. *IEEE Trans. Image Process.* Feb 2006; **15**(2):354–363.
316. Law M, Chung A. A deformable surface model for vascular segmentation. *Proc. Int. Conf. Medical Image Computing and Computer Assisted Intervention*, 2009; 59–67.
317. Delibasis K, Undrill PE, Cameron GG. Designing fourier descriptor-based geometric models for object interpretation in medical images using genetic algorithms. *Comp. Vis. Imag. Under.* 1997; **66**(3):286–300.
318. Vemuri BC, Radisavljevic A. Multiresolution stochastic hybrid shape models with fractal priors. *ACM Trans. Graph.* 1994; **13**(2):177–207.
319. Metaxas D, Koh E, Badler NJ. Multi-level shape representation using global deformations and locally adaptive finite elements. *Int. J. Comput. Vis.* 1997; **25**(1):49–61.
320. Cootes TF, Hill A, Taylor CJ, Graham J. The use of active shape models for locating structures in medical images. *Imag. Vis. Comp.* Jul 1994; **12**(6):355–366.
321. Cootes TF, Edwards GJ, Taylor CJ. Active appearance models. *Proc. European Conf. Comput. Vis.*, 1998; 484–498.
322. Tsechpenakis G, Metaxas DN. CRF-driven implicit deformable model. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2007; 1–8.
323. Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S, *et al.*. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* Jan 2002; **33**(3):341–355.
324. Han X, Hoogeman MS, Levendag PC, Hibbard LS, Teguh DN, Voet P, Cowen AC, Wolf TK. Atlas-based auto-segmentation of head and neck CT images. *Proc. Med. Image Comput. Comput. Assist. Interv.*, 2008; 434–441.
325. Leung KK, Barnes J, Ridgway GR, Bartlett JW, Clarkson MJ, Macdonald K, Schuff N, Fox NC, Ourselin S. Automated cross-sectional and longitudinal hippocampal volume measurement in mild cognitive impairment and Alzheimer's disease. *Neuroimage* Jul 2010; **51**(4):1345–1359.
326. Carmichael OT, Aizenstein HA, Davis SW, Becker JT, Thompson PM, Meltzer CC, Liu Y. Atlas-based hippocampus segmentation in Alzheimer's disease and mild cognitive impairment. *Neuroimage* Oct 2005; **27**(4):979–990.
327. Vemuri BC, Ye J, Chen Y, Leonard CM. Image registration via level-set motion: applications to atlas-based segmentation. *Med. Image Anal.* Mar 2003; **7**(1):1–20.
328. Bauer S, Nolte LP, Reyes M. Segmentation of brain tumor images based on atlas-registration combined with a Markov-Random-Field lesion growth model. *Proc. IEEE Int. Symp. Biomedical Imaging: From Nano to Macro*, 2011; 2018–2021.
329. Fortunati V, Verhaart RF, van der Lijn F, Niessen WJ, Veenland JF, Paulides MM, van Walsum T. Hyperthermia critical tissues automatic segmentation of head and neck CT images using atlas registration and graph cuts. *Proc. IEEE Int. Symp. Biomedical Imaging*, 2012; 1683–1686.
330. Klein A, Hirsch J. Mindboggle: a scatterbrained approach to automate brain labeling. *Neuroimage* Jan 2005; **24**(2):261–280.

331. Heckemann RA, Hajnal JV, Aljabar P, Rueckert D, Hammers A. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *Neuroimage* Oct 2006; **33**(1):115–126.
332. van der Lijn F, den Heijer T, Breteler MMB, Niessen WJ. Hippocampus segmentation in MR images using atlas registration, voxel classification, and graph cuts. *Neuroimage* Dec 2008; **43**(4):708–720.
333. Teng CC, Shapiro LG, Kalet IJ. Head and neck lymph node region delineation with image registration. *BioMedical Engineering OnLine* Jun 2010; **9:30**.
334. Gorthi S, Duay V, Houhou N, Schick MBU, Becker M, Allal AS, Thiran JP. Segmentation of head and neck lymph node regions for radiotherapy planning using active contour-based atlas registration. *IEEE J. Selected Topics in Signal Processing* Feb 2009; **3**(1):135–147.
335. Han X, Hibbard LS, O'Connel N, Willcut V. Automatic segmentation of head and neck CT images by GPU-accelerated multi-atlas fusion. *MIDAS Journal* Jul 2009; :297–304.
336. Gorthi S, Cuadra MB, Schick U, Tercier PA, Allal AS, Thiran JP. Multi-atlas based segmentation of head and neck CT images using active contour framework. *Proc. Med. Image Comput. Comput. Assist. Interv.*, 2010; 313–321.
337. Huttenlocher DP, Klanderman GA, Rucklidge WJ. Comparing images using the Hausdorff distance. *IEEE Trans. Pattern Anal. Mach. Intell.* Sep 1993; **15**(9):850–863.
338. van den Elsen PA, Pol EJD, Viergever MA. Medical image matching - a review with classification. *IEEE Engineering in Medicine and Biology Magazine* 1993; **12**(1):26–39.
339. Maintz JBA, Viergever MA. A survey of medical image registration. *Med. Image Anal.* 1998; **2**(1):1–36.
340. Pluim JPW, Maintz JBA, Viergever MA. Mutual-information-based registration of medical images: A survey. *IEEE Trans. Med. Imag.* Aug 2003; **22**(8):986–1004.
341. Maes F, Vandermeulen D, Suetens P. Medical image registration using mutual information. *Proceedings of the IEEE* Oct 2003; **91**(10):1699–1722.
342. Bookstein FL. Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Trans. Pattern Anal. Mach. Intell.* Jun 1989; **11**(6):567–585.
343. Gueziec A, Ayache N. Smoothing and matching of 3d space curves. *Int. J. Comput. Vis.* 1994; **12**(1):79–104.
344. Feldmar J, Ayache N. Locally affine registration of free-form surfaces. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 1994; 496–501.
345. Lavallée S, Szeliski R. Recovering the position and orientation of free-form objects from image contours using 3D distance maps. *IEEE Trans. Pattern Anal. Mach. Intell.* Apr 1995; **17**(4):378–390.
346. Rueckert D, Sonoda LI, Hayes C, Hill DL, Leach MO, Hawkes DJ. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans. Med. Imag.* Aug 1999; **18**(8):712–721.
347. Tang TWH, Chung ACS. Non-rigid image registration using graph-cuts. *Proc. Med. Image Comput. Comput. Assist. Interv.*, 2007; 916–924.
348. Bajcsy R, Kovacic S. Multiresolution elastic matching. *Comput. Vision, Graphics Image Process.* Apr 1989; **46**(1):1–21.
349. Evans AC, Dai W, Collins L, Neelin P, Marrett S. Warping of a computerized 3D atlas to match brain image volumes for quantitative neuroanatomical and functional analysis. *Proc. SPIE Conf. Medical Imaging V: Image Processing*, vol. 1445, 1991; 236–246.
350. Gee J, Reivich M, Bajcsy R. Elastically deforming 3D atlas to match anatomical brain images. *J. Computer Assisted Tomography* 1993; **17**(2):225–236.
351. Declerck J, Subsol G, Thirion JP, Ayache N. Automatic retrieval of anatomic structures in 3D medical images. *Proc. Int. Conf. Computer Vision, Virtual Reality and Robotics in Medicine*, vol. 905, 1995; 153–162.
352. Subsol G, Thirion JP, Ayache N. A general scheme for automatically building 3D morphometric anatomical atlases: Application to a skull atlas. *Proc. Int. Symp. Medical Robotics and Computer Assisted Surgery*, 1995; 226–233.
353. Davatzikos CA, Prince JL, Bryan RN. Image registration based on boundary mapping. *IEEE Trans. Med. Imag.* 1996; **15**(1):112–115.
354. Sandor S, Leahy R. Towards automatic labelling of the cerebral cortex using a deformable atlas model. *Proc. Int. Conf. Information Processing in Medical Imaging*, vol. 3, 1995; 127–138.
355. So RWK, Chung ACS. Non-rigid image registration by using graph-cuts with mutual information. *Proc. IEEE Int. Conf. Image Processing*, 2010; 4429–4432.
356. Glocker B, Komodakis N, Tziritas G, Navab N, Paragios N. Dense image registration through MRFs and efficient linear programming. *Med. Image Anal.* 2008; **12**(6):731–741.
357. Shekhovtsov A, Kovtun I, Hlaváč V. Efficient MRF deformation model for non-rigid image matching. *Comput. Vis. Image Understand.* 2008; **112**(1):91–99.
358. Arce-Santana ER, Alba A. Image registration using Markov random coefficient and geometric transformation fields. *Pattern Recognit.* 2009; **42**(8):1660–1671.

DOI: 10.1002/cnm

359. Palmer R, Tam G, Xie X, Alock R, Roobottom C. Computing 3D mesh correspondence for aortic root shape modelling. *Proc. 19th Conf. Medical Image Understanding and Analysis*, 2015.
360. Viola P, Wells III WM. Alignment by maximization of mutual information. *Proc. IEEE Int. Conf. Comput. Vis.*, 1995; 16–23.
361. Viola P, Wells III WM. Alignment by maximization of mutual information. *Int. J. Comput. Vis.* 1997; **24**(2):137–154.
362. Moshfeghi M. Elastic matching of multimodality medical images. *CVGIP: Graphical Models and Image Process.* 1991; **53**(3):271–282.
363. Moshfeghi M, Ranganath S, Nawyn K. Three-dimensional elastic matching of volumes. *IEEE Trans. Image Process.* Mar 1994; **3**(2):128–138.
364. Guimond A, Roche A, Ayache N, Meunier J. Three-dimensional multimodal brain warping using the demons algorithm and adaptive intensity corrections. *IEEE Trans. Med. Imag.* 2001; **20**(1):58–69.
365. Thirion JP. Image matching as a diffusion process: An analogy with Maxwell's demons. *Med. Image Anal.* 1998; **2**(3):243–260.
366. Collignon A, Maes F, Delaere D, Vandermeulen D, Suetens P, Marchal G. Automated multimodality medical image registration using information theory. *Proc. Int. Conf. Information Processing in Medical Imaging*, vol. 3, 1995; 263–274.
367. Maes F, Collignon A, Vandermeulen D, Marchal G, Suetens P. Multi-modality image registration by maximization of mutual information. *Proc. Workshop Mathematical Methods in Biomedical Image Analysis*, 1996; 14–22.
368. Studholme C, Hill DLG, Hawkes DJ. Automated 3-D registration of MR and CT images of the head. *Med. Image Anal.* 1996; **1**(2):163–175.
369. Wells III WM, Viola P, Atsumi H, Nakajima S, Kikinis R. Multi-modal volume registration by maximization of mutual information. *Med. Image Anal.* 1996; **1**(1):35–51.
370. Maes F, Collignon A, Vandermeulen D, Marchal G, Suetens P. Multimodality image registration by maximization of mutual information. *IEEE Trans. Med. Imag.* Apr 1997; **16**(2):187–198.
371. Meyer CR, Boes JL, Kim B, Bland PH, Zasadny KR, Kison PV, Koral K, Frey KA, Wahl RL. Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin plate spline-warped geometric deformations. *Med. Image Anal.* 1997; **1**(3):195–206.
372. Studholme C, Hill DLG, Hawkes DJ. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognit.* Jan 1999; **32**(1):71–86.
373. Pluim JPW, Maintz JBA, Viergever MA. Image registration by maximization of combined mutual information and gradient information. *IEEE Trans. Med. Imag.* Aug 2000; **19**(8):809–814.
374. Cheah TC, Shanmugam SA, Mann KAL. Medical image registration by maximizing mutual information based on combination of intensity and gradient information. *Proc. Int. Conf. Biomedical Engineering*, 2012; 368–372.
375. Positano V, Santarelli MF, Landini L, Benassi A. Automatic time sequence alignment in contrast enhanced MRI by maximization of mutual information. *Proc. IEEE Int. Conf. Engineering in Medicine and Biology Society*, vol. 3, 2001; 2407–2410.
376. Bhattacharya M, Das A. Multi resolution medical image registration using maximization of mutual information & optimization by genetic algorithm. *Proc. IEEE Nuclear Science Symp. Conf. Record*, vol. 4, 2007; 2961–2964.
377. Maes F, Vandermeulen D, Suetens P. Comparative evaluation of multiresolution optimization strategies for multimodalitiy image registration by maximization of mutual information. *Med. Image Anal.* 1999; **3**(4):373–386.
378. Zibaeifard M, Rahmati M. An adaptive simulated annealing scheme for multi-modality medical image registration by maximization of mutual information. *Proc. Int. Conf. Signal Processing*, vol. 2, 2006.
379. Palos G, Betrouni N, Coulanges M, Vermandel M, Devlaminck V, Rousseau J. Multimodal matching by maximisation of mutual information and optical flow technique. *Proc. IEEE Int. Conf. Engineering in Medicine and Biology Society*, vol. 1, 2004; 1679–1682.
380. Klein S, Staring M, Pluim JPW. Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines. *IEEE Trans. Image Process.* Dec 2007; **16**(12):2879–2890.
381. Freire L, Godinho F. Registration by maximization of mutual information - a cross validation study. *Proc. IEEE Int. Symp. Bio-Informatics and Biomedical Engineering*, 2000; 322–329.
382. Tsao J. Interpolation artifacts in multimodality image registration based on maximization of mutual information. *IEEE Trans. Med. Imag.* Jul 2003; **22**(7):854–864.
383. Bardinet E, Cohen LD, Ayache N. A parametric deformable model to fit unstructured 3D data. *Comp. Vis. Imag. Under.* 1998; **71**(1):39–54.
384. Nikolova M. Markovian reconstruction using a GNC approach. *IEEE Trans. Image Process.* Sep 1999; **8**(9):1204–1220.

385. Raj A, Singh G, Zabih R. MRF's for MRI's: Bayesian reconstruction of MR images via graph cuts. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 1, 2006; 1061–1068.
386. Benameur S, Mignotte M, Destrempes F, de Guise JA. Three-dimensional biplanar reconstruction of scoliotic rib cage using the estimation of a mixture of probabilistic prior models. *IEEE Trans. Biomed. Eng.* Oct 2005; **52**(10):1713–1728.
387. Cunningham GS, Hanson KM, Battle XL. Three-dimensional reconstructions from low-count SPECT data using deformable models. *Optics Express* Mar 1998; **2**(6):227–236.
388. Lalush DS, Tsui BMW. Block-iterative techniques for fast 4D reconstruction using a priori motion models in gated cardiac SPECT. *Phys. Med. Biol.* 1998; **43**(4):875–886.
389. Mair BA, Gilland DR, Cao Z. Simultaneous motion estimation and image reconstruction from gated data. *Proc. IEEE Int. Symp. Biomedical Imaging*, 2002; 661–664.
390. Yang HF, Choe Y. Cell tracking and segmentation in electron microscopy images using graph cuts. *Proc. IEEE Int. Symp. Biomedical Imaging: From Nano to Macro*, 2009; 306–309.
391. Yang HF, Choe Y. 3D volume extraction of densely packed cells in EM data stack by forward and backward graph cuts. *Proc. IEEE Symp. Computational Intelligence for Multimedia Signal and Vision Processing*, 2009; 47–52.
392. Metaxas DN. *Physics-Based Deformable Models: Applications to Computer Vision, Graphics and Medical Imaging*. Boston: Kluwer Academic Publishers, 1996.
393. Blake A, Isard M. *Active Contours: The Application of Techniques from Graphics, Vision, Control Theory and Statistics to Visual Tracking of Shapes in Motion*. New York: Springer-Verlag, 1998.
394. Ueda N, Mase K. Tracking moving contours using energy-minimizing elastic contour models. *Proc. European Conf. Comput. Vis.*, 1992; 453–457.
395. Gupta A, von Kurowski L, Singh A, Geiger D, Liang CC, Chiu MY, Adler LP, Haacke M, Wilson DL. Cardiac MR image segmentation using deformable models. *Proc. Computers in Cardiology*, 1993; 747–750.
396. Geiger D, Gupta A, Costa LA, Vlontzos J. Dynamic programming for detecting, tracking and matching deformable contours. *IEEE Trans. Pattern Anal. Mach. Intell.* Mar 1995; **17**(3):294–302.
397. Kumar S, Goldgof D. Automatic tracking of SPAMM grid and the estimation of deformation parameters from cardiac MR images. *IEEE Trans. Med. Imag.* 1994; **13**(1):122–132.
398. Amini AA, Curwen RW, Klein AK, Egglin TK, Pollak J, Lee F, Gore JC. Physics based snakes, Kalman snakes, and snake grids for feature localization and tracking in medical images. *Proc. Int. Conf. Information Processing in Medical Imaging*, 1995; 363–364.
399. Kraitchman DL, Young AA, Chang CN, Axel L. Semi-automatic tracking of myocardial motion in MR tagged images. *IEEE Trans. Med. Imag.* 1995; **14**(3):422–432.
400. Park J, Metaxas D, Axel L. Volumetric deformable models with parameter functions: A new approach to the 3D motion analysis of the LV from MRI-SPAMM. *Proc. IEEE Int. Conf. Comput. Vis.*, 1995; 700–705.
401. Young AA, Kraitchman DL, Dougherty L, Axel L. Tracking and finite element analysis of stripe deformation in magnetic resonance tagging. *IEEE Trans. Med. Imag.* 1995; **14**(3):413–421.
402. Park J, Metaxas D, Axel L. Analysis of left ventricular wall motion based on volumetric deformable models and MRI-SPAMM. *Med. Image Anal.* 1996; **1**(1):53–72.
403. Park J, Metaxas D, Young AA, Axel L. Deformable models with parameter functions for cardiac motion analysis from tagged MRI data. *IEEE Trans. Med. Imag.* Jun 1996; **15**(3):278–289.
404. Creswell LL, Wyers SG, Pirolo JS, Perman WH, Vannier MW, Pasque MK. Mathematical modelling of the heart using magnetic resonance imaging. *IEEE Trans. Med. Imag.* 1992; **11**(4):581–589.
405. Young AA, Axel L. Non-rigid wall motion using MR tagging. *Proc. IEEE Int. Conf. Comput. Vis.*, 1992; 399–404.
406. Young AA, Axel L. Three-dimensional motion and deformation of the heart wall: estimation with spatial modulation of magnetization - a model-based approach. *Radiology* 1992; **185**(1):241–247.
407. O'Donnell T, Boult T, Gupta A. Global models with parametric offsets as applied to cardiac motion recovery. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 1996; 293–299.
408. Radeva P, Amini AA, Huang J. Deformable B-solids and implicit snakes for 3D localization and tracking of SPAMM MRI data. *Comput. Vis. Image Understand.* May 1997; **66**(2):163–178.
409. Chandrashekara R, Mohiaddin RH, Rueckert D. Analysis of 3-D myocardial motion in tagged MR images using nonrigid image registration. *IEEE Trans. Med. Imag.* Oct 2004; **23**(10):1245–1250.
410. Shi W, Zhuang X, Wang H, Duckett S, Luong DVN, Tobon-Gomez C, Tung K, Edwards PJ, Rhode KS, Razavi RS, *et al.*. A comprehensive cardiac motion estimation framework using both untagged and 3-D tagged MR images based on nonrigid registration. *IEEE Trans. Med. Imag.* Jun 2012; **31**(6):1263–1275.
411. Young AA. Model tags: direct three-dimensional tracking of heart wall motion from tagged magnetic resonance images. *Med. Image Anal.* 1999; **3**(4):361–372.

412. Chen T, Wang X, Chung S, Metaxas D, Axel L. Automated 3D motion tracking using Gabor filter bank, robust point matching, and deformable models. *IEEE Trans. Med. Imag.* Jan 2010; **29**(1):1–11.
413. Wong WK, Yang B, Liu C, Poignet P. A quasi-spherical triangle-based approach for efficient 3-D soft-tissue motion tracking. *IEEE/ASME Trans. Mechatronics* Oct 2013; **18**(5):1472–1484.
414. Malis E. An efficient unified approach to direct visual tracking of rigid and deformable surfaces. *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, 2007; 2729–2734.
415. Pentland A, Horowitz B. Recovery of nonrigid motion and structure. *IEEE Trans. Pattern Anal. Mach. Intell.* 1991; **13**(7):730–742.
416. Nastar C, Ayache N. Fast segmentation, tracking, and analysis of deformable objects. *Proc. IEEE Int. Conf. Comput. Vis.*, 1993; 275–279.
417. Nastar C, Ayache N. Non-rigid motion analysis in medical images: A physically based approach. *Proc. Int. Conf. Information Processing in Medical Imaging*, 1993; 17–32.
418. Nastar C, Ayache N. Frequency-based nonrigid motion analysis: application to four dimensional medical images. *IEEE Trans. Pattern Anal. Mach. Intell.* 1996; **18**(11):1067–1079.
419. Metaxas D, Terzopoulos D. Shape and nonrigid motion estimation through physics-based synthesis. *IEEE Trans. Pattern Anal. Mach. Intell.* 1993; **15**(6):580–591.
420. Chen CW, Huang TS, Arrott M. Modeling, analysis and visualization of left ventricle shape and motion by hierarchical decomposition. *IEEE Trans. Pattern Anal. Mach. Intell.* Apr 1994; **16**(4):342–356.
421. Bardinet E, Cohen LD, Ayache N. Tracking and motion analysis of the left ventricle with deformable superquadrics. *Med. Image Anal.* 1996; **1**(2):129–149.
422. Amini AA, Owen RL, Anandan P, Duncan JS. Non-rigid motion models for tracking the left-ventricular wall. *Proc. Int. Conf. Information Processing in Medical Imaging*, 1991; 343–357.
423. Duncan JS, Owen RL, Staib LH, Anandan P. Measurement of non-rigid motion using contour shape descriptors. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 1991; 318–324.
424. Amini AA, Duncan JS. Bending and stretching models for LV wall motion analysis from curves and surfaces. *Image Vision Comput.* 1992; **10**(6):418–430.
425. Cohen I, Ayache N, Sulger P. Tracking points on deformable objects using curvature information. *Proc. European Conf. Comput. Vis.*, 1992; 458–466.
426. Mishra SK, Goldgof DB, Huang TS. Motion analysis and epicardial deformation estimation from angiography data. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 1991; 331–336.
427. Kambhamettu C, Goldgof DB. Point correspondence recovery in non-rigid motion. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 1992; 222–227.
428. Huang WC, Goldgof DB. Adaptive-size meshes for rigid and nonrigid shape analysis and synthesis. *IEEE Trans. Pattern Anal. Mach. Intell.* 1993; **15**(6):611–616.
429. Kambhamettu C, Goldgof DB. Curvature-based approach to point correspondence recovery in conformal nonrigid motion. *CVGIP: Image Understanding* 1994; **60**(1):26–43.
430. Donoho DL. Compressed sensing. *IEEE Trans. Infor. Theory* Apr 2006; **52**(4):1289–1306.
431. Candès EJ, Wakin MB. An introduction to compressive sampling. *IEEE Signal Processing Magazine* Mar 2008; **25**(2):21–30.
432. Candès EJ. Compressive sampling. *Proc. IEEE Cong. Mathematicians*, vol. 3, 2006; 1433–1452.
433. Baraniuk RG. Compressive sensing. *IEEE Signal Processing Magazine* Jul 2007; **24**(4):118–121.
434. Block KT, Uecker M, Frahm J. Undersampled radial MRI with multiple coils. iterative image reconstruction using a total variation constraint. *Magnetic Resonance in Medicine* Jun 2007; **57**(6):1086–1098.
435. Lustig M, Donoho D, Pauly JM. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine* Dec 2007; **58**(6):1182–1195.
436. Gamper U, Boesiger P, Kozerke S. Compressed sensing in dynamic MRI. *Magnetic Resonance in Medicine* Feb 2008; **59**(2):365–373.
437. Lustig M, Donoho D, Santos JM, Pauly JM. Compressed sensing MRI. *IEEE Signal Processing Magazine* Mar 2008; **25**(2):72–82.
438. Fowler JE, Mun S, Tramel EW. Block-based compressed sensing of images and video. *Foundations and Trends in Signal Processing* Apr 2012; **4**(4):297–416.
439. Razzaq FA, Mohamed S, Bhatti A, Nahavandi S. Locally sparsified compressive sensing for improved MR image quality. *Proc. IEEE Conf. Systems, Man, and Cybernetics*, 2013; 2163–2167.
440. Chen SS, Donoho DL, Saunders MA. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.* Aug 1998; **20**(1):33–61.
441. Tropp JA, Wright SJ. Computational methods for sparse solution of linear inverse problems. *Proceedings of the IEEE* Jun 2010; **98**(6):948–958.

DOI: 10.1002/cnm