A review of 3D vessel lumen segmentation techniques: Models, features and extraction schemes

David Lesage a,b,*, Elsa D. Angelini b, Isabelle Bloch b, Gareth Funka-Lea a

aSiemens Corporate Research, Imaging and Visualization Dept., Princeton NJ, USA
bInstitut Telecom, Telecom ParisTech, CNRS LTCI, Paris, France

ARTICLE INFO

Article history:
Received 31 March 2008
Received in revised form 17 July 2009
Accepted 24 July 2009
Available online 12 August 2009

Keywords:
Review
Vascular segmentation
Vascular models
Vessel-dedicated features
Segmentation algorithms
3D angiography
Computed tomography angiography (CTA)
Magnetic resonance angiography (MRA)

ABSTRACT

Vascular diseases are among the most important public health problems in developed countries. Given the size and complexity of modern angiographic acquisitions, segmentation is a key step toward the accurate visualization, diagnosis and quantification of vascular pathologies. Despite the tremendous amount of past and ongoing dedicated research, vascular segmentation remains a challenging task. In this paper, we review state-of-the-art literature on vascular segmentation, with a particular focus on 3D contrast-enhanced imaging modalities (MRA and CTA). We structure our analysis along three axes: models, features and extraction schemes. We first detail model-based assumptions on the vessel appearance and geometry which can be embedded in a segmentation approach. We then review the image features that can be extracted to evaluate these models. Finally, we discuss how existing extraction schemes combine model and feature information to perform the segmentation task. Each component (model, feature and extraction scheme) plays a crucial role toward the efficient, robust and accurate segmentation of vessels of interest. Along each axis of study, we discuss the theoretical and practical properties of recent approaches and highlight the most advanced and promising ones.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Vascular diseases are among the most important public health problems in developed countries. For instance, coronary heart disease, resulting from blocked coronary arteries, is currently the first cause of death in the US (Rosamond et al., 2008) and one of the leading causes of death worldwide (WHO, 2008). Such statistics naturally motivate the large amount of research dedicated to vascular imaging.

The segmentation of vascular structures is particularly valuable for diagnosis assistance, treatment and surgery planning. It is indeed a fundamental step for the accurate visualization of vessels from complex datasets and for the quantification of pathologies. Unfortunately, most angiographic clinical routines still rely heavily on manual operations. Given the amount of data generated by modern 3D imaging modalities, such as computed tomography angiography (CTA) and magnetic resonance angiography (MRA), manual segmentation can quickly add up to hours of processing. In this context, automatic and semi-automatic image processing tools aim at easing and speeding up reviewing tasks, reducing the amount of manual interaction and lowering inter-operator variability.

Vascular segmentation is an especially specific and challenging problem. Besides general, acquisition-dependent considerations about contrast, resolution, noise and artifacts, vascular networks can be particularly complex structures. Blood vessels potentially exhibit high variability of size and curvature. Their appearance and geometry can be perturbed by stents, calcifications, aneurysms, and stenoses. Finally, they are often embedded in complex anatomical scenes, surrounded by other organs.

Angiography covers a wide range of anatomical applications (cerebral, retinal, hepatic, peripheral, pulmonary, cardiac . . . ) and modalities (X-ray, CTA, MRA, ultrasound, 2D or 3D . . . ). While not focusing on a particular application, we mainly report recent works on CTA and MRA in 3D. These modern modalities have reached high quality and resolution levels.¹ They allow the capture of highly detailed vascular networks embedded in increasingly large volumes of data. These two challenging aspects (complexity of the target network v.s. large search spaces) motivate the use of advanced, dedicated segmentation schemes.

Our study almost exclusively deals with the most common vascular segmentation task, lumen segmentation. Although most of the general points discussed in this paper also apply to other vascular applications, tasks such as outer wall or thrombus segmentation raise additional, specific challenges.

¹ Cardiac CT angiography, for instance, typically reaches sub-millimetric intra- and inter-slice resolutions with moderate anisotropy, on state-of-the-art multi-slice CT scanners.
In this review, we intend to remain as general as possible. We often do not detail application- and modality-dependent aspects. Instead, we focus on high-level discussions and emphasize general trends and major innovations in state-of-the-art literature. Our goal is to provide the reader with general insight on this particularly prolific field and highlight the key design issues that researchers may encounter in the development of new methods.

Given the wide range of applications considered, direct performance comparisons are also left out. More generally, one can regret a lack of standardly accepted databases and validation criteria for most vascular segmentation applications. Initiatives such as the BrainWeb project (Broche et al., 2006) for brain MRI and the Rotterdam Coronary Artery Algorithm Evaluation Framework (Rotterdam Coronary Artery Algorithm Evaluation Framework, 2008; Schaap et al., 2009a) for coronary CTA segmentation are particularly precious as they allow direct quantitative comparisons between different approaches.

To the best of our knowledge, the currently most general and extensive vascular segmentation review can be found in Kirbas and Quek (2004), with a highly detailed categorization of existing works with respect to their mathematical frameworks. Other existing reviews focus more closely on a limited number of works and applications: CT peripheral angiography in (Feiibel et al., 2001), MR angiography in (Suri et al., 2002a,b), with a first part (Suri et al., 2002a) dedicated to pre-filtering and a second part (Suri et al., 2002b) focusing on skeleton versus non-skeleton extraction schemes. One should additionally mention the review of skeletonization algorithms from Buhler et al. (2002), which also covers visualization techniques, from surface and volume reconstruction, to the generation of meshes and curved planar reconstruction views.

Instead of using a linear categorization of the different frameworks, we discuss existing works according to three high-level axes:

(i) appearance and geometric models (Section 2);
(ii) image features (Section 3);
(iii) extraction schemes (Section 4).

Models correspond to the prior assumptions made on the target vessels, e.g. elongation and hyper-intensity. Features are the vessel-dedicated image measures used to estimate the models on the image, e.g. local intensity curvatures. Finally, the extraction scheme represents the algorithmic core of a vascular segmentation method, i.e., how models, guided by the features, are optimized to perform the segmentation in practice.

A given technique may rely on intricate combinations of models, features and extraction schemes. Table 2 illustrates the decomposition of a selection of works representative of the main trends in the field. Our primary goal is not so much the exhaustive categorization of existing works but the discussion of why proposed combinations are relevant. Each component effects the robustness, accuracy and efficiency of the final algorithm. Throughout this paper, we review the theoretical and practical properties of a large panel of models, features and extraction schemes. Table 2 may be used along with the table of contents (Table 1) to efficiently navigate through the paper according to one’s particular interests.

2. Vascular models

Models, in a broad sense, embed prior information about the target structures. The principal distinction we make here is between appearance (photometric) and geometric information.

2.1. Appearance models

Appearance models express prior knowledge on the luminance properties of the target vascular structures. The appearance of blood vessels in medical acquisitions is of course highly dependent on the imaging modality. For the sake of simplicity, we focus here on contrast-enhanced modalities (CTA and MRA) where vessels are brighter than the surrounding structures.2

2.1.1. Vessel-only models

Knowledge on the theoretical vessel intensity distribution can be derived from the specificities of the imaging protocols. For instance, CT angiographies are generally calibrated so that contrast-enhanced tissues appear in pre-defined intensity ranges. This property was directly applied for vessel segmentation in numerous papers (Boskamp et al., 2004; Sebbe et al., 2003; Wesarg and Firle, 2004; Zahitzen, 1995). Detailed MR-specific appearance modeling aspects are discussed for instance in (Hoogeveen et al., 1998) for time-of-flight (TOF) and phase contrast (PC) acquisitions.

In practice, perfect steadiness of the contrast agent concentration is impossible to achieve and contrast agent inhomogeneity causes appearance variability, even for precisely calibrated modalities such as CTA. Such variability can be handled through the use of statistical distributions, derived from theoretical and/or heuristic knowledge. Various models have been proposed in the literature, such as Gaussian (Florin et al., 2005), Cauchy (Agam and Wu, 2005), Rician (Chung and Noble, 1999) or double sigmoid (Quek and Kirbas, 2001) distributions.

2.1.2. Vessels + background models

When information on surrounding tissues is available, one can use a coupled appearance model, mixing hypotheses on the vascular intensity distribution and surrounding tissues. The simplest example is the assumption that vessels are brighter than their direct neighborhood in contrast-enhanced medical images, an hypothesis exploited in (Florin et al., 2005; Schaap et al., 2007b) among many others. A piece-wise constant model is used in (Tyrrell et al., 2007), with supposedly constant object and background intensities.3 In the work of Hoogeveen et al. (1998) dealing with MR angiography, the background steady state intensity is also modeled as constant. More advanced models are based on statistical mixtures (Chung et al., 2004; Florin et al., 2005; Gan et al., 2004; Hassouna et al., 2003; Manniesing and Niessen, 2004; Wilson and Noble, 1999; Yang et al., 2004). In numerous applications, vessel surroundings contain various structures of highly variable appearance, making the use of explicit background models more difficult. Instead of directly modeling background intensities, the work of Schaap et al. (2007a) on carotid CTA segmentation exploits the assumption that vessels are actually more homogeneous than their surrounding structures.

2.1.3. Noise and point spread function

Besides the appearance of the target vessels and background, noise is another acquisition-dependent aspect that can be modeled, estimated and incorporated as part of an appearance model. Uniform noise estimation was used in (Hassouna et al., 2003; Wilson and Noble, 1999), Rayleigh noise in (Hassouna et al., 2003) and Laplacian noise in (Tyrrell et al., 2007).

The point spread function (PSF) is another important characteristic of the imaging device and scan and reconstruction parameters. It accounts for spatial blurring phenomena observed in the

1 In certain types of angiographic acquisitions, such as black blood MRA, vessels can also appear darker than their surroundings.

2 In the work of Hoogeveen et al. (1998), authors actually work on 3D multiphoton images, where vessels are darker than the background.
image data. Explicit modeling of the PSF (generally as a spatial Gaussian function) was proposed in (Boskamp et al., 2004; La Cruz et al., 2004; Reinhardt et al., 1997; Saba et al., 2003) for instance. More generally, numerous hybrid models (see Section 2.3) and features (see Sections 3.2 and 3.3) include such notions of spatial blurring inherent to the acquisition process. From a theoretical point of view, and although it is generally implicit, one can thus consider that such techniques integrate a model of the underlying PSF.

### 2.2. Geometric models

A key characteristic of blood vessels is their specific shape. Their elongation, in particular, is a valuable prior knowledge which can be encoded through geometric models. Depending on the application, additional knowledge (specific vessel radius ranges and bifurcations for instance) might be incorporated. We will discuss a few examples in Section 2.5. Hereafter, we focus solely on purely geometric models suited for most angiographic applications.

#### 2.2.1. Surface models

Numerous works rely on implicit assumptions on the vessel surface, such as local tubularity. Such assumptions are generally used to model the cross-sectional geometric and appearance properties (see thereafter and Section 2.3). Among techniques exploiting directly a surface model, one can cite (Frangi et al., 1999a, 2000) with B-spline tensor surfaces and the works of (de Bruijne et al., 2003a,b) with active shape models, learned as shape distributions of abdominal aortic aneurysms (Fig. 1). More generally, one may consider that active contour techniques (see Section 4.3) incorporate model-based regularity assumptions on the vessel surface through the evolution forces they employ.

#### 2.2.2. Vessels as 1D curves: the centerline

A vessel can be considered as a primarily 1D structure, defined by its centerline, i.e., the curve centered inside the vessel lumen. In terms of medial-based representations (Blum, 1967), a vessel's skeleton model simply reduces to a 1D curve. Owing to its simplicity and descriptive power, the vascular centerline has become a critically important tool, used in various clinical routines such as visual inspection tasks and stenosis or aneurysm quantification.

A classical assumption is that the centerline exhibits limited curvature. Various parametric models have been proposed to approximate smoothly varying curves: piece-wise linear curves in (Lacoste et al., 2006) and B-splines in (Bouix et al., 2005; Flasque et al., 2001; Frangi et al., 2000; Larralde et al., 2003). In Wong and Chung (2006), cardinal splines were preferred for their intuitive geometrical behavior. Assumptions on the centerline regularity can also be embedded in the extraction process using prediction techniques such as Kalman filtering (Kalman, 1960; Welch and Bishop, 2001) or particle filters (Arulampalam et al., 2002; Doucet et al., 2001). See Sections 4.4 and 4.5, respectively, for further discussions.

#### 2.2.3. Centerline + cross-section models: generalized cylinders

The vessel surface can be modeled with respect to a centerline curve, following generalized cylinder methods (Binford, 1971). A first approach in this vast family of models is to define vessel walls as 2D cross-sectional contours sweeping along the centerline. Classical cross-sectional models include circles, ellipses, star patterns (O’Donnell et al., 1998; Verdonck et al., 1995; Williams et al., 1997) or parametric curves (Li and Ourselin, 2003). As detailed in Section 3.3, such cross-section models help reducing the complexity of the extraction procedure, under the hypothesis that the centerline estimation is reliable. Generalized cylinder models differ in the kind of centerlines and cross-sections they accept and the way they maintain the 3D coherence of the object. The straight homogeneous generalized cylinder model (Ponce et al., 1989) is limited to linear centerlines and cross-sections varying as scaled versions of a reference shape. Numerous extensions have been proposed (Huang and Stockman, 1993; Terzopoulos et al., 1988; Zerroug and Nevatia, 1996). Of particular interest, the extruded generalized cylinder from O’Donnell et al. (1994) handles curved centerlines...
### Table 2
Sets of models, features and extraction schemes decomposed for a selection of illustrative works.

<table>
<thead>
<tr>
<th>Method</th>
<th>Models</th>
<th>Features</th>
<th>Extraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervised wave propagation (Lorenz et al., 2003)</td>
<td>Hyper-intensity</td>
<td>Adaptive threshold</td>
<td>Wave propagation (Section 4.2) via fast-marching (one seed point)</td>
</tr>
<tr>
<td>Curves (Lorigo et al., 2001) (Fig. 17)</td>
<td>1D centerline</td>
<td>Image gradients</td>
<td>Geodesic, mean-curvature flow level-set evolution (Section 4.3)</td>
</tr>
<tr>
<td>Model-fitting tracking (Wörz and Rohr, 2007; Wörz and Rohr, 2008)</td>
<td>Centerline smoothness from Kalman filtering</td>
<td>Medialness cross-section-based feature (Section 3.3, Fig. 9)</td>
<td>Direct centerline tracking from a seed point (Section 4.4):</td>
</tr>
<tr>
<td>Curves (Lorigo et al., 2001) (Fig. 17)</td>
<td>Centerline smoothness from curvature regularization</td>
<td>Corner detector + specific medialness pattern for bifurcation detection</td>
<td>Prediction by adaptive stepping</td>
</tr>
<tr>
<td>Model-fitting tracking (Wörz and Rohr, 2007; Wörz and Rohr, 2008)</td>
<td>Adaptive threshold</td>
<td>Covariance analysis of extracted segments</td>
<td>Correction by local optimization of the medialness response</td>
</tr>
<tr>
<td>Hessian-based Minimal paths (Olabarriaga et al., 2003; Wink et al., 2001; Wink et al., 2004)</td>
<td>Ellipsoid model from Hessian-based features (Fig. 4, Section 2.3)</td>
<td>Local fit of the hybrid model (Section 3.2)</td>
<td>Bifurcation detection</td>
</tr>
<tr>
<td>4D minimal paths (Li and Yezi, 2006) (Fig. 25)</td>
<td>4D paths: (centerline + radius)</td>
<td>2 Intensity-based features proposed:</td>
<td>Multi-hypothesis, stochastic Bayesian tracking via particle filtering (Section 4.5)</td>
</tr>
<tr>
<td>Particle filtering (Florin et al., 2005; Florin et al., 2006) (Fig. 26, Section 4.5)</td>
<td>1D centerline curve</td>
<td>Cross-sectional features:</td>
<td>Direct tracking from seed point Item–Clustering of particles for bifurcation detection</td>
</tr>
<tr>
<td>Markov marked point processes (Lacoste et al., 2006) (Section 4.5.2, Fig. 27)</td>
<td>&quot;Quality Candy&quot; segment model:</td>
<td>Contrast operator on linear segments</td>
<td>Markov marked point process:</td>
</tr>
</tbody>
</table>

- Hyper-intensity
- Locally tubular geometry
- Centerline smoothness
- Branching frequency
- 1D centerline from curvature regularization
- Varying, limited vessel radius
- 1D centerline curve
- Ellipsoid model from Hessian-based features
- 1D centerline curve
- Cross-section geometry model:
- Elliptic in (Florin et al., 2005)
- Circular shortest paths in (Florin et al., 2006)
- Gaussian mixture appearance
- Local hyper-intensity/contrast
- Non-linear variations of geometric and appearance parameters
- Polyline refined tree model:
- Connectivity/influence between segments
- Local contour contrast
- Multi-hypothesis, stochastic Bayesian tracking via particle filtering (Section 4.5) |
- Wave propagation (Section 4.2) via fast-marching (one seed point)
- Model-fitting tracking (Wörz and Rohr, 2007; Wörz and Rohr, 2008)
- Hessian-based Minimal paths (Olabarriaga et al., 2003; Wink et al., 2001; Wink et al., 2004)
- 4D minimal paths (Li and Yezi, 2006) (Fig. 25)
- Particle filtering (Florin et al., 2005; Florin et al., 2006) (Fig. 26, Section 4.5) | Markov marked point process: |

- Local fit of the hybrid model (Section 3.2) | Multi-hypothesis, stochastic Bayesian tracking via particle filtering (Section 4.5) | Combination of stochastic processes: |

- Medialness cross-section-based feature (Section 3.3, Fig. 9) | Corner detector + specific medialness pattern for bifurcation detection | Initial segment process |

- Corner detector + specific medialness pattern for bifurcation detection | Correction by local optimization of the medialness response | Refined polyline hierarchical process |

- Direct centerline tracking from a seed point (Section 4.4): |

- 2 Intensity-based features proposed: | Multi-hypothesis, stochastic Bayesian tracking via particle filtering (Section 4.5) | Fully-automatic, random initialization |
and varying cross-sections while avoiding torsion artifacts common to previous models.

Extending these ideas, the works from Frangi et al. (1999a, 2000) model the centerline as a 1D B-spline curve and use a 3D B-spline (NURBS) surface evolution to fit the vessel wall along the curve (see Fig. 2). Similarly, a 3D active surface under axial constraints is proposed in (Montagnat, 1999) and a level-set evolution from an initial centerline is used in (van Bemmel et al., 2003). These methods come with schemes to optimize the surface with respect to the data (see Section 4.3) and yield finer fitting than classical generalized cylinders based solely on 2D cross-sections.

2.2.4. A contrario models

An original approach consists in describing a vessel a contrario, by defining what it should not be. Examples include the ball measure from Nain et al. (2004) and the non-local measures used in (Rochery et al., 2005a, 2006). These soft shape priors are used in active contour frameworks to penalize local widening, maintaining the shape elongation a contrario (see Section 4.3 and Figs. 18 and 19). As it is generally inextricable to model every non-vessel structure in the image, we expect such an approach to be limited to relatively low-level, soft geometric constraints.

2.3. Hybrid models: appearance + geometry

Hybrid models combine appearance and geometric information by incorporating assumptions on the spatial appearance of a vessel.

2.3.1. Radial intensity profiles

It is generally too complex to define the intensity distribution as a function of space for highly variable shapes such as generalized cylinders (see Section 2.2). The geometric part of hybrid models is thus usually limited to a local approximation of the vessel. The spatial intensity distribution is then modeled as radial variations in the corresponding cross-sections, introducing the notion of intensity profiles.

Mathematically appealing profiles include parabolic profiles (Steger, 1996) and 3D Gaussian line profiles (Gong et al., 2003; Krissian et al., 2000; Rohr and Wörz, 2006; Shikata et al., 2004), for which saliency criteria (Steger, 1996) and estimation bounds (Rohr and Wörz, 2006) were analytically derived. In practice, vessels (especially large ones) often do not have rounded but rather plateau-like profiles, which are not adequately described by the aforementioned models. The simplest model accounting for that observation is the bar-like profile (Boldak et al., 2003; Koller et al., 1995; Steger, 1998), which corresponds to a local cylinder with homogeneous intensity. Of particular interest are the analytical developments in (Steger, 1998) and the use of an asymmetric profile for heterogeneous surroundings, with analytical derivation of the corresponding bias correction.

Bar-like profiles do not account for phenomena such as spatial blur and partial volume effect. A first approach to integrate such effects in a bar-like model is the 3D Phi line from Rohr and Wörz (2006), based on the Phi Gaussian error function. Another popular model is the bar-like convolved profile, where the cross-section radial appearance is defined as a plateau convolved with a Gaussian kernel (Krissian et al., 2000; Wörz and Rohr, 2004, 2007, 2008) (see Fig. 3). Estimation bounds are provided in (Wörz and Rohr, 2008). In such models, spatial blur is considered as acquisition-dependent. As previously stated in Section 2.1, such effects are manifestations of the actual point spread function (PSF) of the imaging device. In this regard, the bar-like convolved model can be seen as a PSF appearance model combined with a 3D cylinder geometric model. Estimation of such models can thus be considered as an inverse, deconvolution problem aiming at estimating the vessel’s location, direction and radius given the geometric and PSF models (see also Section 3.2).

Finally, the work of Hoogeveen et al. (1998) focuses on the fundamental limits to the accuracy of vessel diameter estimations from MR time-of-flight (TOF) and phase contrast (PC) acquisitions. They base their numerical study on 3D appearance models taking into account acquisition-dependent factors such as flow velocity, radio-frequency pulses, relaxation times and flip angles. These models assume a 3D tubular geometry, with laminar flow along the principal axis. Different intensity profiles are proposed for TOF and PC acquisitions.

2.3.2. Vessels as ridges

Another important body of literature deals with vessels as local ridges of the image hyper-surface (with intensities considered as an extra dimension added to the spatial ones). Such models can be viewed as extensions of Blum’s medial representation (Blum, 1967) from binary to gray-level images, aiming at exploiting both
edge- and region-based information. Although they generally do not integrate explicit appearance models, ridge models qualify as hybrid models as they correlate spatial and intensity information by specifying the desired mathematical properties of ridge locations. Various ridge definitions have been proposed (Eberly et al., 1994), among which height ridges, based on the idea of estimating intensity maxima in special directions. 3D, multiscale extensions of height ridges were exploited in several successful vascular-dedicated techniques (Aylward et al., 1996; Aylward and Bullitt, 2002; Fridman et al., 2003). We refer the readers to Eberly et al. (1994) for more details on the mathematical framework of the large family of ridge models.

2.3.3. Template-based shape spaces

Frangi et al. (1998) introduce a shape space based on second-order intensity variations (Hessian matrix analysis), further refined in (Lin and May, 2003). Shape prototypes are defined through relative variations of Hessian eigenvalues. Vessels are matched to bright, elongated ellipsoids (see Fig. 4). Hessian-based features to detect such shapes in the image are discussed in Section 3.2. A similar shape space is defined in (Hernandez-Hoyos, 2002) based on inertia moments.

Related to generalized cylinders (Section 2.2), superellipsoids (Tyrrell et al., 2007) locally describe the vessel geometry in combination with a constant intensity appearance model (see Fig. 5). Instead of being embedded in an image tensor (Hessian or inertia matrix), the parameterization (3D pose and appearance) is made explicit. Similarly, the linear templates from Friman et al. (2008), based on a Gaussian line appearance model, and the linear contrast operators from Lacoste et al. (2005, 2006) are constructed from explicitly parameterized 3D cylinders.

2.4. Models of bifurcations and anomalies

Besides regular vessel segments, bifurcations and medically critical anomalies such as calcifications, aneurysms, stents and stenoses can also benefit from dedicated modeling. The scattered apparition of bifurcations and anomalies in angiographies is generally hard to predict, so precise modeling can considerably help their robust detection.

2.4.1. Bifurcations

Theoretical geometric bifurcation models were first developed in (Murray, 1926), including relationships between branching angles and vessel widths based on physiological optimality conditions (see Fig. 6). This model has been evaluated experimentally in various publications such as (Fanucci et al., 1988; Zamir and Brown, 1982, 1983) and further extended in (Oka and Nakai, 1987). Accurate bifurcation models are also proposed in (Antiga and Steinman, 2004; Flaaris et al., 2004), but are not directly suited for segmentation purposes as they require a pre-extracted surface mesh and centerline for their optimization. Simpler geometric models can be found in (Agam and Wu, 2005; Agam et al., 2005), where possible bifurcations are categorized as Y or X junctions.
(with three and four principal directions, respectively). In 
Hernandez-Hoyos (2002), bifurcations are defined a contrario as deviations from cylindrical segments in terms of moments of inertia. In practice, most works do not model bifurcations explicitly. Numerous techniques focus solely on single branch segmentation, while other employ dedicated detection schemes, based notably on topology analysis.

2.4.2. Calcifications and stents

In contrast-enhanced CTA data, calcifications and stents appear as hyper intense structures and are classically handled with specific appearance models, typically resulting in the definition of higher-bound thresholds (Boldak et al., 2003; Wesarg and Firle, 2004).

2.4.3. Stenoses and aneurysms

Stenoses are generally modeled implicitly as local and sudden radius decrease (Frangi et al., 2000; Hernandez-Hoyos et al., 2000; Sato et al., 1998a; Sun et al., 1995; Verdonck et al., 1996). Among explicit stenosis models useful for segmentation, one can mention the double cone model (Frangi et al., 1998; Hernandez-Hoyos, 2002; Lin and May, 2003) used in (Lin and May, 2003) to derive a specialized detector based on Hessian analysis.

Aneurysms, due to their high shape variability, are generally not explicitly modeled. A counter-example is the work from de Bruijne et al. (2003b), where a model of abdominal aortic aneurysms is learned from a database (see Fig. 1). More generally, special effort is often put into making methods robust to aneurysms (Lin and May, 2003; Wink et al., 2000a), so that they can be detected a posteriori as local deviations from the vessel model. Experiments from the literature suggest local radius and volume increase as possible indicators of most prominent aneurysms. Refined morphological criteria are employed in (Kawata et al., 1995).

2.5. Problem-specific models

Application-, organ-specific information may also be incorporated to ease the segmentation task. We already mentioned the work of de Bruijne et al. (2003b) on abdominal aortic aneurysms (see Fig. 1) as an example of application-specific learned model. In Florin et al. (2004), authors use the heart surface to constrain a ray-casting/profile analysis to likely locations of coronaries in cardiac CT angiographic data (see Fig. 12). In Shikata et al. (2004), a lung mask is used for pulmonary artery segmentation. In the relatively rare applicative contexts where reliable statistical data is available, atlases can be employed to guide vascular segmentation, as done in (Passat et al., 2006) for brain vasculature.

2.6. General considerations on vessel models

Numerous factors influence practical model choices. Prior knowledge on the appearance of the vessels of interest is often specific to the target modality. On the other hand, the adequacy of a geometric model depends on the application (size, regularity and complexity of the target vessels), but also on the eventual clinical goals. Simple centerline models are generally sufficient for visual screening tasks, but more complete, accurate representations such as generalized cylinders are required for the quantification of vascular pathologies.

A single segmentation algorithm may naturally rely on several complementary models. For instance, a hybrid model can be used for local geometry and appearance, with generalized cylinders controlling the geometric coherence of the result at a higher level. Models can also be combined in coarse-to-fine, multi-layer designs. A simple centerline model can guide a coarse, but fast and robust delineation of the vessels of interest. Such a rough delineation can then be used to efficiently initialize and constrain a refined segmentation process, governed by a more complex model. Such an approach was for instance exploited in (Frangi et al., 1999a, 2000) and benefit to the overall robustness and computational efficiency of the segmentation scheme.

3. Features

Hereafter, we review existing vessel-specific features, i.e., the practical detectors and filters used to evaluate vascular models on the data. For each technique, we describe which kind of information is exploited and how it is extracted. Besides their theoretical background, we discuss their robustness, accuracy, discriminative power and computational efficiency.

Vessel-dedicated features are commonly designed by combining and adapting simple, well-spread image processing tools, from basic image intensities to first and second-order edge-based detectors (e.g. gradients, Canny detector, Laplacian zero-crossings). The reader may refer to general and medical image processing reviews and handbooks (Ayache, 1995; Duncan and Ayache, 2000; McInerney and Terzopoulos, 1996; Sonka et al., 1996; Sonka and Fitzpatrick, 2000) for extensive overviews of these classical techniques.

Features of interest incorporate vessel-specific prior knowledge (typically derived from an underlying model) in their designs, e.g. assumptions on the elongation and the direction, on the scale and circularity of vascular cross-sections. Scale and directionality are two aspects essential to multiscale line filters, from which numerous vessel features are derived. We refer the reader to Lindeberg (1994), ter HaarRomeny (2003) for comprehensive overviews of line detection in scale-space frameworks.

3.1. Local isotropic features

A few vessel-dedicated features from the literature solely focus on estimating the location and/or scale of target vessels. We refer to such features as isotropic, as they do not exploit assumptions on the directionality of the vessels.

In contrast-enhanced modalities, the simplest assumption one can exploit is that vessels are brighter than their immediate surroundings. In Szymczak et al. (2005), authors rely on local, scale-persistent robust intensity maxima to select a set of candidate points, expected to lie mostly inside the vessels. Similarly, local intensity maxima on 2D slices are selected as seed points for the tracking scheme of Lee et al. (2007).

In Aylward et al. (1996), authors implement a 3D, multiscale ridge detector (see Section 2.3). This medialness operator is based...
on multiscale Laplacians of Gaussians, assuming that vascular intensities are locally maximum. Extensions of this work (Aylward and Bullitt, 2002; Fridman et al., 2003; Furst, 1999), explicitly using vessel directionality, are presented in Section 3.3.

Another noticeable technique in this category is the multiscale spherical flux measure first introduced in (Vasilevskiy and Siddiqi, 2002). This feature approximates the divergence of image gradients by measuring the gradient flux through the boundaries of multiscale spheres. Its response is maximal at central vessel locations and corresponding scale. It was proved to be able to detect even low-contrast and narrow vessels. An efficient Fourier-based implementation of the spherical flux feature was recently proposed in (Law and Chung, 2009).

3.2. 3D Local geometry features

Besides hyper-intensity, a core characteristic of vessels is their specific local geometry. Techniques have thus been proposed to locally analyze image data in 3D, with the aim of distinguishing hyper-intense, tubular-like patterns.

3.2.1. Derivative features

A popular approach in the detection of vascular patterns is the use of second-order derivative information (principal curvatures of image intensities) to characterize the local image geometry. The core assumption of second-order derivative methods is that vessels can be distinguished by a locally prominent low curvature orientation (the vessel direction) and a plane of high intensity curvature (the cross-sectional plane). The Hessian matrix is the most common tool to capture such information. To the authors’ knowledge, it was first used for vascular segmentation in (Koller et al., 1995) to estimate vessel orientations thanks to Hessian eigenvectors. Following this seminal idea, a fair amount of work was dedicated to multiscale Hessian-based filters for vessel enhancement (Frangi et al., 1998; Li et al., 2003; Lin and May, 2003; Lorenz et al., 1997; Sato et al., 1998b; Shikata et al., 2004). These filters rely on Hessian eigenvalues to discriminate between plane-, blob- and tubular-like structures, according to prototype-based shape models (see Section 2.3 and Fig. 4). The vesselness filter from Frangi et al. (1998) (see Fig. 7) has been extensively used in practice, owing to its intuitive geometric formulation. The Weingarten matrix is a less popular alternative to the Hessian matrix (Armande et al., 1996; Prinet et al., 1996).

Hessian- and Weingarten-based filters may suffer from sensitivity to local deformations. Aneurysms, stenoses and bifurcations are generally reported to produce false negative responses as such situations sensibly deviate from the assumption of a single strong orientation. Noise and nearby non-vascular structures may also perturb the response of such filters.

These techniques commonly rely on multiscale frameworks to cope with different vessel sizes. Hessian and Weingarten matrices are classically computed in a Gaussian linear scale-space, through convolutions by Gaussian derivatives of different standard deviations (Lindeberg, 1994). Scale selection is performed by keeping the maximum response over multiple scales. The extraction of large vessels requires large standard deviations which induce substantial spatial blur. The response of these filters can thus be perturbed by the presence of other hyper-intense structures in the immediate surroundings of the target vessel. Finally, depending on the range of selected scales, these filters may yield substantial computational costs.

The recent work of Bauer and Bischof (2008b), Bauer and Bischof (2008a) is an interesting alternative to multiscale Hessian-based filters. The Hessian matrix can be viewed as the Jacobian of the gradient vector field. Instead of relying on the Gaussian scale-space gradient vector field, Bauer and Bischof (2008b) propose to use an other vector field, obtained from gradient vector flow (GVF) diffusion (Xu and Prince, 1998). The GVF scheme regularizes and anisotropically diffuses gradients from the object boundaries. This vector flow vanishes at medial locations, i.e., on vessel centerlines. Bauer and Bischof (2008b) apply the vesselness criteria from Frangi et al., 1998 on the local Jacobian of the GVF field to derive their enhancement filter. This approach overcomes most of the shortcomings of Gaussian scale-space techniques. It is mono-scale and less sensitive to perturbations from structures adjacent to the target vessels. To alleviate the high computational cost of GVF diffusion, Bauer and Bischof (2008b) use a GPU implementation.

Instead of analyzing second-order derivatives of the image, one can alternatively exploit the spatial covariance of image gradients vectors. A first subgroup of such methods focuses on the local distribution of gradient vectors through the analysis of the structure tensor matrix (covariance matrix of gradient vectors) (Weickert, 1999) to exhibit clusters of gradient vectors. Works from Agam and Wu (2005) and Agam et al. (2005) share the ability of detecting potentially more than one principal direction. They can thus discriminate between simple vascular segments, nodules and bifurcations. In Agam and Wu (2005), estimations are performed through Expectation–Maximization of a probabilistic mixture model, while iterative component analysis is preferred in (Agam et al., 2005). In terms of scale adaptation, (Agam and Wu, 2005; Agam et al., 2005) implement heuristic rules to select the size of the window of interest. In Agam et al. (2005), authors favorably compare their methods to Hessian-based filters (Frangi et al., 1998; Li et al., 2003; Shikata et al., 2004), reporting smaller mean square localization errors and a better behavior at bifurcations.

Another recent approach exploiting the distribution of gradient vectors is the optimally oriented flux from Law and Chung (2008). Similarly to the spherical flux (Vasilevskiy and Siddiqi, 2002; Law and Chung, 2009) discussed in Section 3.1, the optimally oriented flux relies on the measure of gradient flux through the boundary of local spheres. The key difference lies in the exploitation of the vessel directionality through the estimation of an optimal gradient projection axis. Law and Chung (2008) develop a Fourier-based

![Fig. 7. Hessian-based multiscale filter from Frangi et al. (1998). Application to the enhancement of aortoiliac MRA. From left to right: original image, four increasing scale response images, combined in the final enhanced image. Illustration based on material from Frangi et al. (1998).](image-url)
implementation and discuss the advantages of optimally oriented flux over Hessian-based measures, including more accurate and stable responses and higher robustness to disturbances from adjacent structures.

Closely related, several oriented features use first order derivative information along 2D cross-sectional circular patterns (Fridman et al., 2003; Fridman, 2004; Gulsun and Tek, 2008; Krissian et al., 2000; Wink et al., 2000a). These filters are explicitly oriented, i.e., the direction to be tested is an input parameter. Such cross-sectional features are discussed in Section 3.3.

3.2.2 Integrative features

As an alternative to derivative methods, some authors proposed to use integrative measures such as second-order inertia moments of image intensities (Boldak et al., 2003; Hernandez-Hoyos et al., 2000, 2006; Hernandez-Hoyos, 2002; Larralde et al., 2003; Reuze et al., 1993). These filters are theoretically less sensitive to noise while providing similar information on local geometry. Vessel detection filters can be designed similarly to Hessian-based ones, with ratios of the eigenvalues of the inertia matrix (Hernandez-Hoyos et al., 2000, 2006; Hernandez-Hoyos, 2002). Contrary to second-order derivative methods, integrative techniques do not enjoy a well-defined scale-space theoretical framework. Most works use a fixed, single scale. Heuristics to adapt its value are proposed in (Hernandez-Hoyos et al., 2000, 2006; Hernandez-Hoyos, 2002).

In the binary case, the inertia matrix reduces to the spatial covariance matrix of the binary shape. In Lorenz et al. (2003), such a covariance matrix is used to iteratively check pre-extracted vessel segments and heuristically adapt segmentation parameters. The scale-independent elongation criterion of Wilkinson and Westenberg (2001) also uses spatial covariance for vessel enhancement. This criterion is applied to the connected components given by the grayscale levels of the image, in the context of morphological connected filters (Serra and Salembier, 1993).

3.2.3 Local fit of 3D hybrid models

In Section 2.3, we described how radial appearance profiles can be used to model intensity variations locally, in tubular regions of interest. If such a hybrid model can be expressed (or approximated) in a closed form, partial derivatives with respect to the parameters can be derived analytically. Then, the hybrid model can be directly fitted to the local data, typically with non-linear least square optimization such as Levenberg–Marquardt schemes. This method was successfully applied for Gaussian profiles in (Friman et al., 2008) and bar-like convolved models in (La Cruz et al., 2004; Wörz and Rohr, 2004, 2007, 2008). Goodness of fit at convergence can be used as a measure to discriminate between vascular and non-vascular areas. Measures of fit significance, evaluated with Student t-tests, can additionally be derived (Friman et al., 2008).

Although not relying on an explicit hybrid model, the framework of Lacoste et al. (2005, 2006) also uses linearly oriented elements as local contrast operators. Student t-tests quantify the significance of contrast values. No local optimization is performed, as the segmentation process is globally driven by a Markov marked point process (see Section 4.5.2). In Tyrrell et al. (2007), local superellipsoids (see Fig. 5) rely on an estimation of the local contrast using a piece-wise constant appearance model (constant object on a constant background). Local optimization is performed through log-likelihood maximization. Significance is measured through generalized likelihood ratio tests, in the context of robust M-estimators. Authors emphasize the robustness of this estimation scheme to nearby structures such as adjacent vessels.

3.2.4 A contrario features

The previous features rely on the general principle of maximizing the fit of a vascular model. Another interesting approach is to detect vessels a contrario. In Section 2.2, we already mentioned the ball measure from Nain et al. (2004) and the non-local measures from Rochery et al. (2005a, 2006). These purely geometric features are used to guide active contour optimizations by penalizing deviations from tubular-like shapes (see Section 4.3 and Figs. 18 and 19).

Another feature in this category is the top-hat filter. This morphological operator detects local intensity peaks through their suppression by morphological opening. Its general formulation can be tailored to exploit the thinness and elongation of vascular structures by using linear oriented patterns as structuring elements. When oriented perpendicularly to a vessel in the image, such oriented patterns will cause its suppression by opening, hence its detection by the top-hat filter. The directional top-hat sum technique from Zana and Klein (2001) is illustrated in Fig. 8.

3.3 2D cross-sectional features

Some authors focused their analysis on the cross-section geometry, testing hypotheses such as its compactness and regularity. Cross-sectional features can be viewed as special cases of 3D local geometry features from Section 3.2. They have the advantage of downscaling the analysis from three to two dimensions. The orientation to be tested are considered as an input parameter. Directional information is commonly estimated from features such as Hessian eigenvalues (see Section 3.2) or given by the extraction scheme (see Section 4). A single orientation estimation may be sensitive to noise and local perturbations. A robust but potentially costly alternative is to test exhaustively a discrete set of possible directions.

![Fig. 8. Oriented top-hat from Zana and Klein (2001). (a) Opening with linear structuring elements. The 2D vessel is illustrated as a 3D crest line (intensity elevation map). Three structuring elements are given. The structuring element perpendicular to the vessel will give the best top-hat response, as it can not be contained in the vessel and will remove it from the opening. The two other directional openings will leave the vessel nearly unaltered. (b) Left, original image (2D retinography). Right, sum of oriented top-hats. Illustrations based on material from Zana and Klein (2001).](image-url)
In Behrens et al. (2001), a randomized 2D ellipse Hough transform is used to detect 2D cross-sections on edge maps. While computationally efficient, such a technique does not provide very accurate localization and heavily relies on the underlying edge detector. Following similar ideas, approximate circle fitting, based on maxima selection of a multiscale medial filter, is proposed in (Rueckert et al., 1997).

In Florin et al. (2005), the so-called ribbon measure compares the mean intensity inside a 2D ellipse with its immediate surroundings. This contrast-based feature reaches local maxima when the ellipse is well aligned with the vessel cross-section. Parameters are optimized in a particle filter framework (see Section 4.5).

### 3.3.1. Cross-sectional circular medialness measures

In the context of ridge tracking (see Section 2.3), the work from Aylward and Bullitt (2002) extends their original medialness feature (Aylward et al., 1996) by exploiting the vessel’s directionality, through a 2D convolution kernel measuring the contrast on circular cross-sections. Direction is estimated through the Hessian eigenvectors and a new scale optimization scheme is proposed.

The work of Koller et al. (1995) on steerable derivative filters introduced the seminal ideas shared by many cross-sectional medialness features. Koller et al. (1995) base their initial 2D analysis on the optimal line detector from Canny (1986) convolved with a bar of a variable width. They rely on two shifted, oriented Gaussian derivative kernels to detect parallel vessel walls. A multiscale scheme handles varying vessel widths. Authors discuss the theoretical properties of their filters, such as the response to different models of vessel profiles (Section 2.3). Of particular interest, Koller et al. (1995) propose a non-linear combination (through a min operator) of the shifted kernels to reduce step-edge false responses. The extension to 3D was the first vascular filter, to our knowledge, to use the Hessian matrix to estimate the local orientation. Only two directions are tested in the cross-sectional plane, whose responses are again combined with a min operator.

Very similar ideas are developed in Poli and Valli, 1996, in 2D, with a particular focus on kernel separability for computational efficiency. As noted by the authors, these linear filters are subject to false positive responses at step edges. A basic intensity comparison between “left” and “right” positions is proposed to alleviate that shortcoming. Another closely related approach is the so-called ellipsoidal scale-space filter from Suri et al. (2002c).

Inspired by Aylward and Bullitt (2002), Koller et al. (1995) and Furst (1999), the work from Fridman et al. (2003), Fridman (2004) develops a formulation of the core framework (Pizer et al., 1998) dedicated to vascular structures. A medial atom response is obtained by convolving the image with directional Gaussian derivative kernels located at the end of equi-angular radial spokes (see Fig. 9). Given a location, an orientation and a radius, the core feature measures the fit of the image edges with a circular cross-section model. Bifurcations are handled thanks to an additional edge detector coupled with specific patterns of the core response. Similarly, the adaptive filter from Krissian et al. (2000) uses a sum of oriented gradient responses along a circle. In Krissian et al. (2000), special attention is given to the analytical and experimental study of the filter with respect to scale estimation and response to different appearance radial profiles. Orientation is classically given by Hessian eigenvectors.

### 3.3.2. Ray-casting features

The idea of ray-casting has received much interest for the characterization of 2D vessel cross-sections (Gulsun and Tek, 2008; Niessen et al., 1999; Tek et al., 2001; Verdonck et al., 1995; Wesarg and Firle, 2004; Wink et al., 1998, 2000a). Given a point p in the vessel and a direction d defining the cross-sectional plane, ray-casting techniques aim at detecting the vessel wall along rays thrown equi-angularly from p. Detected contour points can then be used to evaluate the centerness of a point inside a vessel. Existing ray-casting features are closely related to the medialness measures from the previous paragraph. Instead of relying on a circular or elliptic model, they only assume a compact, regular cross-sectional contour.

As depicted in Fig. 10, ray-casting reduces the detection problem to a 1D analysis along the rays, yielding in general high levels of computational performance. Intensity (Wesarg and Firle, 2004) or gradient thresholding (Tek et al., 2001; Wink et al., 2000a) is classically used to select candidate contour points. Robustness is ensured by casting and averaging numerous rays. Additionally, constraints on the spatial coherence of neighboring contour points can be enforced through simple heuristics, or, more globally, thanks to an optimization by dynamic programming (Verdonck et al., 1995). The center of mass of candidate contour points is often employed to measure the centerness of a location (Niessen et al., 1999; Verdonck et al., 1995; Wesarg and Firle, 2004; Wink et al., 1998, 2000a). A notable alternative is the use of mean-shift filtering in (Tek et al., 2001).

In Gulsun and Tek (2008), ray-casting is used to design a multiscale medialness operator, similarly to other filters described at the beginning of this section. Responses along rays are summed in a distance-wise manner, following a circular cross-section model. A local normalization scheme is employed to convey contrast invariance. This new medialness measure makes use of the sign of the gradient along a ray to lessen the influence of nearby hyper-intense structures. Gradient polarity along rays was also exploited in (Wesarg and Firle, 2004; Wink et al., 2000a) to handle hyper-intense anomalies such as calcifications.

### 3.3.3. Accurate extraction of 2D contours

Techniques performing an accurate segmentation of 2D cross-sectional contours can also be considered as special cases of

---

Fig. 9. Core medial atom from Fridman et al. (2003). The length of equi-angular spokes corresponds to the radius of the vessel being tested. Weighting functions (directional Gaussian derivatives), at the end of each spoke, are summed to obtain the filter response. Illustration based on material from Fridman et al. (2003).

Fig. 10. Ray-casting technique from Wink et al. (2000a). From a point in the vessel, rays are cast in a 2D cross-section and gradient maxima are selected as contour points. Note the robustness to calcifications. Application to CT angiography. Illustration based on material from Wink et al. (2000a).
cross-section features. Similarly to most ray-based techniques, such methods do not rely on fixed geometric models. Instead, their optimization schemes integrate soft compactness and regularity constraints on the cross-sectional contours. Classical techniques in that category include 2D parametric active contours, as used in (Hernandez-Hoyos, 2002; Li and Ourselin, 2003) (see Fig. 28). In Rueckert et al. (1997), active contours are coupled with stochastic relaxation to overcome local minima issues. One can also highlight the family of circular dynamic-programming methods, such as the 2D star pattern from Geiger et al. (1995), used in (O’Donnell et al., 1998; Verdonck et al., 1995), the circular shortest path from Appleton and Sun (2003) used in (Florin et al., 2006) and the mean graph cycle optimization used in (Gulsun and Tek, 2006, 2008). These techniques benefit from optimal and relatively efficient optimization schemes.

3.4. Features for bifurcations and anomalies

3.4.1. Bifurcations

Bifurcations can be detected by the emergence of several principal directions, as exploited by the clustering method in (Agam et al., 2005). In Hernandez-Hoyos (2002), branching point candidates are selected through an inertia moment-based heuristic. These ideas are further refined in (Carrillo et al., 2007), with the analysis of connected components on a local sphere surface through a moment-based clustering algorithm. This notion of binary connected components was also exploited in (Flasque et al., 2001), with a local parallelepipeded ROI encompassing the extracted segment. In Fridman et al. (2003), authors use responses of a corner detector (Blom, 1991) combined with a bifurcation-specific adaptation of their core filter. If the features provide a width estimate, local increase of this value can be used as a clue for the detection of bifurcations (Aylward and Bullitt, 2002). Most other branch handling schemes are based on a posteriori topological analysis of segmented results and are dependent on the extraction method (Section 4).

3.4.2. Stenoses and aneurysms

Lin and May (2003) propose a stenosis detector based on Hessian analysis. Stenosis detection is more generally performed a posteriori by quantifying radius decrease (Frangi et al., 2000; Hernandez-Hoyos et al., 2000; Sato et al., 1998a; Sun et al., 1995; Verdonck et al., 1996), by blood flow simulation (Deschamps et al., 2004) or by 3D curvature analysis (Kawata et al., 1995, 1996).

Similarly to stenoses, aneurysms can be detected a posteriori as local radius increases, or by 3D curvature analysis (Kawata et al., 1995, 1996). Although none has been proposed to the authors’ knowledge, one could probably design Hessian-based aneurysm detectors, as hinted in (Sato et al., 1998b). Other techniques dedicated to aneurysms include learned shape models (de Bruijne et al., 2003a,b) and volume-based estimation (Hernandez et al., 2003).

3.4.3. Calcifications

Calcifications are usually handled simply as hyper-intense areas inside the vessel. In Wesarg and Firle (2004), calcifications are separated from the vessel lumen by dual thresholding. In Wink et al. (2000a), calcifications are detected as high positive gradient along concentric rays in 2D cross-sections.

3.5. General considerations on vessel features

Existing vascular features differ naturally in terms of accuracy, robustness and computational efficiency. Another important consideration to take into account is the amount of information that a given feature extracts from the image. For instance, popular Hessian-based detectors require only test locations to be specified and provide both direction and scale estimations. However, these low-level cues are potentially corrupted by local image perturbations. 2D cross-sectional features generally provide robust scale estimations, but require both location and direction as inputs. Input parameters can be viewed as defining the search space to be explored by the overall extraction process. Specialized features, with several parameters, yield more complex optimization schemes. On the other hand, the extraction process can, in general, incorporate higher-level, model-based information to compensate the unreliability of low-level, feature-based estimations. The balance between low-level, high-level, data-based and model-based information is a critical and intricate design issue.

4. Extraction schemes

In this section, we review the optimization schemes proposed in the literature to effectively perform vessel segmentation. We discuss possible algorithmic designs, relying on models from Section 2 and exploiting features from Section 3. First, we briefly review some pre-processing and pre-segmentation approaches to simplify image content, improve image quality and obtain an approximate localization of the vessels (Section 4.1). Such preliminary results may serve as inputs for vessel-dedicated extraction schemes, which we separate in three general classes:

- region-growing algorithms (4.2);
- active-contours (4.3);
- centerline-based approaches (4.4).

Additionally, we highlight two promising stochastic techniques recently applied to vascular segmentation: particle filtering and Markov marked point processes (Section 4.5). We finally discuss various post-processing techniques to refine segmentation results in Section 4.6. The reader may refer to Fig. 11 as a guide for this section’s layout.

4.1. Pre-processing and pre-segmentation

4.1.1. Pre-processing

To simplify the tremendous amount of information embedded in 3D angiograms, one can first consider generic techniques such as down-sampling and quantization (Tschirren et al., 2005). One can also improve the image quality through filtering. Numerous features from Section 3.2 were originally proposed as vessel enhancement filters (see Fig. 7). This includes derivative filters from Koller et al. (1995), Poli and Valli (1996), structure tensor analysis (Agam and Wu, 2005; Agam et al., 2005), mathematical morphology filters (Wilkinson and Westenberg, 2001; Zana and Klein, 2001), popular Hessian-based filters (Frangi et al., 1998; Krissian et al., 2000; Li et al., 2003; Lin and May, 2003; Lorenz et al., 1997; Sato et al., 1998b; Shikata et al., 2004) and flux-based filters (Law and Chung, 2008, 2009). Another filtering technique not previously mentioned is the vessel enhancement method from Westin et al. (2001), which relies on a direction estimation through frequency-based quadrature filters to perform locally adaptive low-pass and high-pass filtering.

More generally, a fair amount of research effort can be found in vessel-dedicated anisotropic diffusion schemes (Gerig et al., 1992; Krissian et al., 1997; Manningsing et al., 2006; Meijering et al., 2001; Orkisz et al., 1997; Weickert, 1999) for angiography enhancement. These regularization schemes cope with the challenge of reducing noise levels while preserving thin and weakly contrasted vessels. Comprehensive overviews and evaluations of state-of-the-art approaches are proposed in (Meijering et al., 2001 and Suri et al., 2002b). Diffusion is guided by the local anisotropy of the data,
classically derived from the Hessian matrix (Krissian et al., 1997; Manniesing et al., 2006) or structure tensors (Weickert, 1999). Vectorial gradient flux was also used for anisotropic diffusion and was demonstrated to be successful at preserving thin, low-contrast vessels (Krissian, 2002). Finally, the recent work from Tankyevych et al. (2008) can also be considered as a diffusion scheme in a broad sense. Morphological filters are applied with respect to a tracking process, guided by Hessian-based direction information. Tankyevych et al. (2008) focus more particularly on the reconnection of noisy tubular structures.

4.1.2. Pre-segmentation

Most of the extraction algorithms of the next sections require some rough, prior localization of the target vessels. Regions of interest (ROIs) can first be selected given prior anatomical knowledge. Examples include the use of a mask of the heart for coronary segmentation in (Florin et al., 2004) (Fig. 12), a mask of the lungs for pulmonary vessels in (Shikata et al., 2004) and a probabilistic atlas for cerebral vasculature in (Passat et al., 2006). ROIs can also be obtained by thresholding a potential map. Image intensities were directly exploited in (Wu et al., 2004). A more advanced approach is to derive mixture probability maps from an appearance model (Section 2.1), estimated by Expectation–Maximization (Chung et al., 2004; Gan et al., 2004; Hassouna et al., 2003; Wilson and Noble, 1999; Yang et al., 2004) (see Fig. 13), or from vessel-enhanced images (Agam and Wu, 2005; Agam et al., 2005; Shikata et al., 2004).

Schemes have also been designed to select only a sparse set of candidate points. Robust maxima are exploited in (Szymczak et al., 2005) (Fig. 29). In Florin et al., 2004, intensity peaks are detected along 1D rays (Fig. 12). Similarly, local maxima are used in (Lee et al., 2007) as seed points for a tracking algorithm (see Section 4.4).
Most of these pre-segmentation methods are based on pixel-wise processes which may be prone to fragmented outputs, false positive and false negative detections. Correction schemes, ensuring spatial coherence or particular topological properties, are discussed in Section 4.6.

4.2. Region-growing approaches

From seed points or regions located inside a vessel, region-growing approaches incrementally segment an object by recruiting neighboring voxels based on some inclusion criteria. Some of these schemes require seed points to be located precisely, at the root of the vascular tree for instance. Inputs are then generally provided manually. A key to the popularity of region-growing schemes, besides their simplicity, is certainly their computational efficiency. They fall in the category of greedy algorithms, employ low-level, simple inclusion rules, and explore datasets only sparsely, a critical advantage for large 3D datasets.

4.2.1. Classical region-growing

In Boskamp et al. (2004), an initial segmentation is obtained via region-growing based on a simple intensity threshold (see Fig. 14). Such a criterion is sensitive to noise and contrast agent inhomogeneities. As classical region-growing are voxel-wise processes, they are also prone to false negative (holes) and false positive (leakage) issues. In Metz et al. (2007), growth-limiting criteria are proposed to lessen risks of leakage. Alternatively, dual object and background competitive region-growing is proposed in (Yi and Ra, 2003). The morphological dilation-based growing scheme from Masutani et al. (1998) also aims at reducing such topological issues.

Algorithmic enhancements of classical region-growing implementations can be found for instance in (Wan et al., 2000, 2002) with symmetric region-growing. The corkscrew algorithm from Wesarg and Firle (2004) can also be considered as a particular variant of region-growing ideas. On a locally thresholded image, linear search paths, cycling standard 3D directions, are followed until the vessel wall is found. Data are explored very sparsely, only along 1D discrete rays. Such a particular search strategy makes an interesting bridge with direct centerline tracking techniques discussed in Section 4.4.

4.2.2. Wave propagation

Wave propagation techniques, also known as front propagations, evolve a well-formed interface inside the vessel (Fig. 15). Their advantage over classical region-growing approaches is the spatially coherent propagation they enforce, which eases advanced dynamical and topological analysis. Discrete waves were proposed in (Bruijns, 2001; Cai et al., 2006; Kirbas and Quek, 2002, 2003; Quek and Kirbas, 2001; Tuchschmid et al., 2005; Zahlten, 1995). These can be seen as ordered region-growing schemes, where candidates for inclusion are ordered according to fitness criteria with local schemes to correct the geometry of the growing front (Zahlten, 1995). A more accurate yet still computationally efficient alternative is the fast-marching algorithm (Adalsteinsson and Sethian, 1995; Sethian, 1996a; Tsitsiklis, 1995). Initially derived

Fig. 12. Ray-casting candidate selection from Florin et al. (2004). Rays are cast from the center of the heart and analyzed for intensity peaks. Selected points form clouds likely to be located in cardiac vessels. Illustration based on material from Florin et al. (2004).

Fig. 13. EM-based segmentation from Yang et al. (2004). Left, original image (coronary CTA). Right, maximum a posteriori image for blood class, including vessels and heart chambers. Anisotropic smoothing is used to ensure spatial coherence. Illustration based on material from Yang et al. (2004).

Fig. 14. Region-growing for vascular segmentation. From left to right: starting from a seed point, a region-growing algorithm successively adds neighboring voxels fulfilling inclusion criteria (e.g., an intensity threshold).

Fig. 15. Wave propagation for vascular segmentation. From the seed point, a lens-like interface, also known as wave or front, is propagated along the vessel.
as a special case derivation of general level-set approaches (see Section 4.3), it can be used as an ordered propagation scheme. Applied on the image grid and given an adequate vessel-dedicated potential, it visits voxel candidates according to an estimation of their geodesic distance from seed points. Another closely related use of fast-marching schemes is the optimization of minimal paths (see Section 4.4). The various vessel-dedicated fast-marching-based growing methods vary in their definitions of the wave speed function: uniform Euclidean speed and binary inclusion criterion in (Lorenz et al., 2003); weighted geodesic speed in (Sebbe et al., 2003); dynamically adapted speed in (Manniesing and Niessen, 2004). In addition, stabilization schemes, used to correct the orientation of the front in high curvature areas, can be found in (Yan and Zhuang, 2003).

4.2.3. On adaptivity

A spatially coherent propagation yields important benefits, such as easier topology analysis. The connectivity of the propagating interface can be exploited to detect bifurcations (Lorenz et al., 2003; Metz et al., 2007; Sebbe et al., 2003; Sekiguchi et al., 2005; Zahlten, 1995) (see Fig. 15). Branching handling can be further eased by a local approximation of the vessel centerline, which can be derived from the successive front centers (Lorenz et al., 2003; Tuchschmidt et al., 2005), by skeletonization of the local binary mask (Manniesing and Niessen, 2004; Tschirren et al., 2005) (see Section 4.6), or by path backtracking (Kirbas and Quek, 2002, 2003; Quek and Kirbas, 2001; Stefanick and Sonka, 2001). The latter approach is closely related to minimal path methods from Section 4.4.

As greedy processes relying on low-level binary criteria, region-growing techniques heavily depend on the robustness of the underlying features. To cope with potential under- and over-segmentation issues, some works have implemented online schemes to dynamically adjust parameters such as inclusion thresholds (Larralde et al., 2003; Lorenz et al., 2003; Manniesing and Niessen, 2004; Masutani et al., 1998; Sekiguchi et al., 2005; Tschirren et al., 2005; Yi and Ra, 2003). Such adaptive schemes exploit the iterative, coherent nature of the propagation process. In Lorenz et al. (2003), Sekiguchi et al. (2005) for instance, it is stressed that adapting parameters independently for each vascular branch is a key factor for accurate and robust extraction. Although typically of a heuristic nature, dynamically adaptive criteria are simple, powerful examples of active collaborations between the extraction process, the features and the underlying models.

4.3. Active contours

Active contours constitute a popular class of image segmentation techniques that evolve an interface through different forces: external forces, derived from the image, and internal, model-based forces, constraining the contour geometry and its regularity (McInerney and Terzopoulos, 1996). In Section 3.3, we already mentioned the special case of active contours for 2D cross-section segmentation as an advanced example of a cross-section feature. For direct 3D extraction of vascular structures however, classical implementations of active contours, e.g. using curvature-based regularization, are usually unfit as they prevent the capture of thin, elongated surfaces. Hereafter, we present refinements specifically targeting vessel-like objects.

4.3.1. Parametric active contours (snakes)

Parametric active contours (Kass et al., 1987) rely on an explicit Lagrangian formulation of a contour evolution. Their advantages are their computational efficiency and ease of implementation in 2D. In 3D however, their parameterization becomes more complex.

4.3.2. Implicit active contours

Implicit active contours have become very popular for image segmentation, especially through the development of level-set techniques. They rely on an Eulerian formulation of contour evolution through partial derivative equations. The contour is embedded as the zero level of a higher dimension (level-set) function. The level-set framework does not suffer from parameterization problems inherent to explicit techniques and topology changes are handled implicitly. Such advantages come at an additional computational cost and, in practice, special algorithmic care has to be taken to ensure convergence. For a more detailed discussion of these
techniques, please refer to reference books such as Angelini et al., 2005, Sethian (1996b, 1999).

Curves (Lorigo et al., 2001) is a vessel-dedicated codimension-two level-set scheme. Its core theoretical principle is to evolve a 1D curve on a 3D domain. In practice, the algorithm of Lorigo et al. (2001) evolves a width-limited surface, called the $C^1$-level-set. A new energy term constrains the lowest curvature of this surface, assumed to be the vessel’s principal direction. This approach has proved to be well adapted to vascular segmentation, as depicted in Fig. 17, and has been subsequently used in (Luboz et al., 2005).

Higher order active contours (Rochery et al., 2005a, 2006) introduce new geometric constraints on the contours, expressed as non-local quadratic forces modeling intricate contour interactions. In particular, these forces penalize local widening of the contour and repulse proximate branches. Examples of evolution patterns are given in Fig. 18. The method was applied in 2D on satellite images for road extraction and on 2D angiograms. Its extension to 3D raises considerable discretization and computational issues. A notable contribution is the choice of phase fields over classical level-set implementations (Rochery et al., 2005b) for neutral initializations. The ball measure from Nain et al. (2004) follows similar ideas. It was proposed as a soft shape prior for classical level-set evolution, penalizing local widening of the contour (see Fig. 19).

Some recent developments have focused on variational formulations of flux maximization as a better suited alternative to curvature-based regularization. Flux maximization aims at aligning the surface normals to a data-based vector field (typically, the gradient vector field). Such an idea was first applied to vascular segmentation in (Vasilevski and Siddiqi, 2002) with the derivation of a multiscale gradient flow (see Fig. 20), and was later extended to multiscale Hessian-based flow in (Descoteaux et al., 2004). Following similar ideas, the edge alignment flow from Kimmel and Bruckstein (2003) was later combined in (Holtzman-Gazit et al., 2006) with Chan and Vese’s minimal partition model (Chan and Vese, 2001), in a hierarchical segmentation framework. The use of the minimal partition model notably ensures global optimality of the energy minimization. The so-called capillary force from Yan and Kassim (2005) is another edge alignment term proposed for level-set evolutions. Authors report increased accuracy over the Curves technique from Lorigo et al. (2001). All these methods exploit vectorial information and share variational formulations closely related to anisotropic filtering techniques from Section 4.1.

Active contours enjoy a flexible, general framework that allows the integration of various internal and external forces. Practical issues may come from classical gradient-based Euler–Lagrange optimization schemes, which converge to local minima of the energetic formulation. The optimization landscape may contain numerous spurious minima, so that in practice, some active contour techniques require initializations close to the final result. The use of vessel-dedicated features and forces, coupled with robust optimization processes such as Kimmel and Bruckstein (2003) help

Fig. 17. Co-dimension 2 level-set: Curves (Lorigo et al., 2001). From left to right, original image (CT lung scan of bronchi), result obtained by classical co-dimension 1 level-set, Curves segmentation and super-imposition of the results. Extra structures obtained by Curves are depicted in light gray. Illustration based on material from Lorigo et al. (2001).

Fig. 18. Higher order active contours from Rochery et al. (2005a). Examples of geometric evolutions without data convection terms, with quadratic forces constraining the contour to tubular-like patterns. Illustrations for different sets of geometry-controlling parameters, based on material from Rochery et al. (2005a).

Fig. 19. Ball measure from Nain et al. (2004). (a) Ball measure on a 2D example: darker pixels indicate higher measurements. (b) Application to coronary CTA. From left to right: evolution of a level-set extraction using the ball measure. The ball measure penalizes widening and progressively disconnects the artery from the heart chamber into which the initial segmentation leaked. Illustrations based on material from Nain et al. (2004).
lessening such requirements. Initialization constraints can be further relaxed by exploiting the implicit handling of topology changes in level-set techniques. For instance, one can use isolated seeds along the vessels that will eventually reconnect (see Fig. 17). False positive seed regions are also tolerated to a certain point as they are expected to disappear under contour regularization. Implicit, thus uncontrolled topology changes may as well become a liability, causing segments to spuriously merge or disconnect. Following the same observations, branchings are usually tolerated, but not explicitly detected.

4.4. Centerline-based methods

The aforementioned region-growing and active contour methods (Sections 4.2 and 4.3) aim at explicitly and directly detecting vessel contours, generally using low-level operations such as pixel-wise inclusion criteria. Centerline-based techniques focus on directly extracting the vessel centerline (see Fig. 21). Arguably, such methods are not true lumen segmentation techniques, as complete volume information is generally not part of the final result. As emphasized in (Aylward and Bullitt, 2002), such techniques seek additional robustness by relying on higher-level information, such as the localization of the center of the vessel, the estimation of its direction and scale. In most cases, a rough volume segmentation can thus be obtained directly by enriching the centerline with the underlying scale information. The extracted centerline can be used to efficiently constrain an accurate segmentation of the contours (see Section 4.6).

Fig. 20. Flux-based level-set from Vasilevskiy and Siddiqi (2002). From top to bottom, left to right, original image (detail of 3D computed rotational angiography of the head), initialization, successive evolution steps until convergence. Notice the capture of thin vessels (arrows). Illustration based on material from Vasilevskiy and Siddiqi (2002).

Fig. 21. Direct centerline tracking (Wink et al., 2000a). Results samples on the ascending aorta and iliac branches from CT angiography. Illustration based on material from Wink et al. (2000a).
prediction and correction steps, following schemes classically used for the statistical estimation of dynamic processes (Kitagawa and Gersch, 1996; West and Harrison, 1997).

The prediction of successive centerline positions is generally performed with respect to estimations of the vessel direction, which may come from model-based and/or image-based sources: geometric models (Section 2.2) and/or image features (Section 3.2). The simplest model-based prediction scheme is to reuse directly the previous direction (Tek et al., 2001; Tyrrell et al., 2007). For increased robustness, filtering of the position history can be employed (Flasque et al., 2001; Frangi et al., 1999b, 2000; Wink et al., 2000a). Kalman filtering, which is a theoretically optimal prediction and correction scheme under Gaussianity and linearity assumptions, has also been applied to vascular tracking (Gong et al., 2003; Wörz and Rohr, 2004, 2007, 2008). The local orientation of the vessel can be estimated from the data thanks to local geometry features (Section 3.2). Hessian-based estimation was used in (Aylward and Bullitt, 2002), moments of inertia in (Boldak et al., 2003; Hernandez-Hoyos et al., 2000; Hernandez-Hoyos, 2002; Hernandez-Hoyos et al., 2006; Larralde et al., 2003; Reuze et al., 1993), and gradient vector distribution in (Agam and Wu, 2005; Agam et al., 2005).

To ensure robust tracking, predicted positions are generally corrected, i.e., recentered. Sphere-based geometric recentering schemes have been proposed in (Carrillo et al., 2005, 2007; Hernandez-Hoyos et al., 2006), relying on local thresholding and inertia moments, respectively. Recentering can also be performed by a local optimization of medialness features. Simple examples are found in (Verdonck et al., 1995; Wesarg and Firle, 2004), where the predicted centerline position is recentered as the center of mass of cross-sectional contour points detected along 1D rays. 2D cross-sectional features (Section 3.3) are particularly popular in this context (Aylward and Bullitt, 2002; Fridman et al., 2003; Fridman, 2004; Niessen et al., 1999; Tek et al., 2001; Verdonck et al., 1995; Wink et al., 1998, 2000a). Applied on a local neighborhood (Aylward and Bullitt, 2002; Fridman et al., 2003; Wink et al., 2000a), recentering schemes have proved to be robust to predictions lying outside the vessel lumen (Wink et al., 2000a). Correction can be further refined with 2D active contours performing an accurate segmentation of the cross-section (Lee et al., 2007) (see Section 4.6). As an alternative to 2D cross-sectional recentering, correction by local optimization of 3D models was used in (Friman et al., 2008; Tyrrell et al., 2007; Wörz and Rohr, 2008).

Direct tracking schemes explore the image very sparsely, reaching levels of time and memory performance suitable for large 3D datasets. The explicit embedding of geometric models generally results in improved robustness over previously mentioned methods. Nonetheless, they remain local optimization processes prone to premature stopping in presence of anomalies such as stenoses and aneurysms. In practice, some of these methods require a fair amount of user interactivity (Aylward et al., 1996; Larralde et al., 2003). A solution to further improve robustness is the use of multi-hypotheses frameworks such as stochastic particle filters (see Section 4.5). A deterministic alternative was recently proposed in (Friman et al., 2008), with a multi-hypothesis algorithm tracking in parallel the successive local maxima of a 3D template fitting process. Substantial gain in robustness is obtained with the increase of the depth of the search tree.

Most centerline tracking techniques follow one branch at a time, relying on manual reseeding to extract a complete tree. To automatize bifurcation handling, some authors have proposed to perform the segmentation of the lumen locally and rely on topological criteria, similarly to aforementioned region-growing and wave propagation schemes (Section 4.2). For instance, the techniques from Carrillo et al. (2005, 2007), Flasque et al. (2001) perform the analysis of connected components on local spheres and parallellepipeds. Due to the use of local binary criteria, these approaches may suffer from the same robustness issues as region-growing techniques. In Aylward and Bullitt (2002), Fridman et al. (2003), online branching detection is performed through specific ridge-based features, without the need for a binary segmentation. An a posteriori branch detector, reportedly preliminary and not entirely satisfactory, is proposed in (Hernandez-Hoyos, 2002). In Wink et al. (2000a), the tracking process is seeded from the smallest branches of the vascular tree and a specific pattern of a medialness filter is exploited to handle branchings. If a robust termination criterion is available, complete tree tracing can theoretically be achieved with automatic reseeding (Tyrrell et al., 2007).

4.4.2. Model-based optimization between two points

If start and end points are available for a given vascular segment, the centerline extraction problem can be advantageously constrained. A first approach to solve such a problem consists in optimizing a centerline model (see Section 2.2) with fixed boundary conditions. B-spline (Frangi et al., 1999a, 2000) (see Fig. 2) and cardinal spline (Wong and Chung, 2006) deformable models have been used in that context. Such methods balance data and model-fitting through the parameterization of their energetic formulation. Optimization is classically performed by gradient descent schemes, which may be prone to local minima issues. In practice, such effects are mitigated by the additional robustness brought by the model.

4.4.3. Minimal path techniques

Certain energetic models for two-point centerline extraction can be formulated as the minimization of a cumulative, monotonic cost metric integrated along the centerline path. The continuous formulation of minimal path optimization can be shown to correspond to a special case of geodesic active contours (Cohen and Kimmel, 1997). Efficient dynamic programming schemes are known to solve such problems globally. Minimal path approaches are particularly popular for centerline extraction. The cost metric is designed to favor centerline locations, which can be done for instance through the reciprocal of a vessel-dedicated feature. Global optimality is a key motivation for such techniques, resulting in good robustness even in cases of corrupted data or severe anomalies (see Fig. 22).

Minimal path techniques differ notably in the numerical optimization schemes they employ. Discrete, \( L_1 \) path optimization can be carried out with Dijkstra-like Dijkstra (1959) graph-based schemes (Gulsun and Tek, 2008; Olabariaga et al., 2003; Wink et al., 2004, 2001, 2000b). Such algorithms have generally the advantage of high computational efficiency, but can be shown to introduce metrication errors due to their discrete nature (Deschamps, 2001). The fast-marching algorithm (Adalsteinsson and Sethian, 1995; Sethian, 1996a; Tsitsiklis, 1995) is an optimization scheme which can be shown to be consistent with the continuous formulation of the minimal path problem (Deschamps, 2001; Cohen and Kimmel, 1997). Fast-marching optimization approximates the Euclidian \( L_1 \) cumulative cost, yielding sub-voxel accurate paths. A counterpart, compared to discrete \( L_1 \) techniques, is a slight increase in computational cost due to more complex update schemes. Among other works, fast-marching optimization has been applied to vascular centerline extraction in (Avants and

---

4 Such assumptions may not hold in practice because of sudden changes in curvature along the vessel. Particle filters are alternative schemes dealing with non-linear, non-Gaussian processes (see Section 4.5).

5 A popular feature in this context is the Hessian-based vesselness measure from Frangi et al. (1998), used for instance in (Olabariaga et al., 2003; Wink et al., 2001, 2004).
Williams, 2000; Deschamps and Cohen, 2001; Law and Heng, 2000; Lin and May, 2003; Young et al., 2001).

Closely related to the aforementioned minimal path approaches, fuzzy connectedness techniques (Udupa et al., 2002) (see also Section 4.6) correspond to the optimization of the $L_1$ norm of a connectivity cost along discrete paths. In other words, the total cost of a path is that of its worst edge (lowest degree of connectivity). The ordered region-growing algorithm of Choyke et al. (2001), Yim et al. (2000, 2003) follows the same principles. The ordered growing pattern defines an acyclic graph composed of the minimal $L_1$ paths from the seed point to all the other points (see Fig. 23). Vessels can then be extracted by interactively specifying end points.\footnote{In Choyke et al. (2001), Yim et al. (2000, 2003), authors focus on automatically cleaning the graph using trimming techniques (see also Section 4.6).} $L_1$ path optimization can be performed efficiently, but it values paths in a non-cumulative fashion, which can result in the extraction of fairly irregular paths in practice.

Although algorithmically efficient, minimal path methods can induce the exploration of a large portion of the search space in order to extract a single centerline. Bidirectional propagation, i.e., computing paths from the start and end points simultaneously, considerably lowers computational costs (Deschamps and Cohen, 2001; Olabarriaga et al., 2003; Wink et al., 2000b). It is used in (Wink et al., 2000b) conjunction with heuristic $A^*$ search to further reduce the search space. Another possible refinement is the freezing scheme from Deschamps and Cohen (2002a), which consists in prevents paths with too high cumulative costs from being propagated further.\footnote{The freezing scheme is used in (Deschamps and Cohen, 2002a) to extract the vessel surface. This corresponds to its use as a binary stopping criterion in a wave propagation scheme (see Section 4.2).} Lin and May (2003) combines the freezing scheme with anisotropic versions of the fast-marching algorithm to greatly reduce the amount of exploration (see Fig. 24).

Classical issues with minimal path techniques include the extraction of eccentric or even erroneous shortcut paths (see Li and Yezzi (2006, 2007) and Fig. 25). These practical problems come...
principally from the cumulative nature of the optimization. They are alleviated in practice by the use of adequate, discriminative cost metrics. Strategies to reduce centering problems a posteriori can be found in (Deschamps and Cohen, 2001, 2002a).

Classical vessel-dedicated minimal path algorithms focus solely on extracting the centerline curve. Notable exceptions are the works from Wink et al. (2004), Li and Yezzi (2006, 2007), Benmansour et al. (2009), Benmansour and Cohen (2009), which share the core idea of incorporating an additional dimension corresponding to the vessel radius. The spatial position of the centerline and its associated radius are then optimized conjointly. This idea was first applied to 2D vascular segmentation in (Wink et al., 2004), using the multiscale feature from Frangi et al. (1998). It was extended to 3D in the 4D minimal path technique from Li and Yezzi (2006, 2007). The authors proposed two potentials expressed on multiscale spheres, based respectively on similarity and contrast criteria. The methods from Wink et al. (2004) and Li and Yezzi (2006, 2007) are implemented as relatively straightforward modifications of the original Dijkstra and fast-marching algorithms, respectively. As important benefits, resulting paths are better centered and the vessel surface can be directly approximated by the envelope of the optimized circles/spheres (see Fig. 25).

Fig. 24. Refinements for fast-marching minimal paths (Lin and May, 2003). (a) Original image (start point in red), (b) Classical isotropic propagation, (c) Isotropic fast-marching with freezing from Deschamps and Cohen (2002a), (d) Orientation map yielding directional information, (e) Anisotropic fast-marching without freezing, (f) Anisotropic fast-marching with freezing. Illustration based on material from Lin and May (2003).

Besides radius optimization, some works focused on exploiting directional information in minimal path techniques. We already mentioned the development of anisotropic fast-marching flows in Lin and May (2003) (Fig. 24) favoring propagation in certain orientations. This idea was recently coupled with multiscale optimization in (Benmansour et al., 2009; Benmansour and Cohen, 2009).
with a particular focus on the design of an adequate multiscale anisotropic metric and on the numerical optimization scheme. Authors in Benmansour et al. (2009), Benmansour and Cohen (2009) rely on the multiscale optimally oriented flux from Law and Chung (2008) as an underlying feature. Another approach to directional minimal paths is to consider the orientation domain as an additional dimension to be optimized, similarly to the aforementioned multiscale techniques. To the authors' knowledge, the application of this idea to angiographic data was first discussed in (Deschamps, 2001). The recent work of Pèchaud et al. (2009) combines multiscale and orientation optimization by propagating 4D (space + scale + orientation) paths on 2D images. In particular, the authors demonstrate the capacity of their method to disambiguate overlap issues in 2D angiograms. The principal drawback of such advanced techniques is their relatively high computational cost, potentially prohibitive for 3D applications.

Minimal path techniques are commonly employed in interactive frameworks, requiring the definition of start and end points for each target vessel. Some works have proposed the definition of termination criteria to automatically stop the path propagation and relax the need for end points. Such criteria can for instance be devised through heuristic thresholds on the underlying features (Gulsun and Tek, 2008). An alternative proposal is the analysis of pioneer points, defined by the first paths crossing a certain geodesic distance (Wink et al., 2001). Finally, Deschamps and Cohen (2002a) proposed to exploit the increase rate of the geodesic distance to stop the propagation. Given a stopping criterion, complex vascular trees can be extracted a posteriori, at convergence, by backtracking multiple minimal paths to a single proximal seed point (Gulsun and Tek, 2008). A related approach can be found in the geodesic voting technique of (Rouchdy and Cohen, 2008), recently applied to 3D vascular trees in (Mille and Cohen, 2009). This method uses a measure of the density of minimal paths running through a given location to identify the main segments of the vascular tree. It relies on the observation that minimal paths from different end points converge and concentrate in highly anisotropic areas corresponding to true vessels. This technique can actually be applied to extract vascular trees from fully optimized action maps, without the need for a stopping criterion. Geodesic voting then serves as an a posteriori pruning criterion, keeping only locations of high path densities (Mille and Cohen, 2009).

4.5. Stochastic frameworks: particle filters and markov marked point processes

In this section, we highlight two stochastic frameworks which were recently successfully adapted to vascular segmentation: particle filtering and Markov marked point processes. In both cases, seminal works showed especially promising results in terms of extraction robustness. We emphasize the overall elegance and versatility of these schemes. They benefit from well-established Bayesian theoretical frameworks and enable high-level designs, seamlessly integrating model- and data-based information.

4.5.1. Particle filters

Sequential Monte-Carlo techniques, also known as particle filtering, are recursive Bayesian estimation schemes which can be used, among other applications, to track the posterior probability distribution of a dynamic process (Arulampalam et al., 2002; Doucet et al., 2001). We already mentioned Kalman filtering as the optimal tracking scheme for linear Gaussian state-space problems. Unfortunately, such simple assumptions generally do not hold for vessel characteristics (sudden scale variations, direction changes, presence of bifurcations). Instead of relying on a deterministic estimation scheme, particle filters handle non-linear processes through stochastic, population-based sampling schemes. Probabilistic tracking, the work from Schaar et al. (2007a) introduces a new appearance model for vessels with heterogeneous surroundings, combined with a tubular geometric model. A subsequent work from the same authors (Schaap et al., 2007b) details algorithmic optimizations coping with redundant computations. The work from Lesage et al. (2008) proposes a new sampling scheme, based on a constrained medial-based geometric model. It discusses the efficient construction of a discrete representation of the posterior density. Such a situation is handled with clustering and repopulation methods. Illustration based on material from Florin et al. (2005).

Fig. 26. Particle filtering from Florin et al. (2005). Application to the tracking of coronary arteries from cardiac CTA. Top row, particle mean states are overlaid on the original image. Bottom row, particle distributions split at bifurcations. Such a situation is handled with clustering and repopulation methods. Illustration based on material from Florin et al. (2005).
4.5.2. Markov marked point processes

Throughout this document, we emphasized the importance of using high level information on the vasculature geometry and appearance. Unfortunately, advanced models often come with numerous parameters, inducing very large search spaces and complex optimization landscapes. Some of techniques previously detailed, such as particle filtering, are able to cope with high-dimensional spaces. On the other hand, they still rely on local optimization schemes and, generally, on user interaction for their initialization. A promising alternative is found in the fully-automatic, globally optimal framework offered by Markov marked point processes.

Markov marked point processes were used in (Lacoste et al., 2006) to extract road networks from remote sensing images. This seminal work was later adapted to the segmentation of coronary arteries from 2D angiograms in (Lacoste et al., 2006) (see Fig. 27). Vessels are modeled locally as piece-wise linear segments of varying locations, lengths, widths and orientations. Lacoste et al. (2006) use simple contrast measures defined on these segments. The extraction process proceeds as follows. A population of segments is initialized randomly on the image and is evolved through a reversible jump Markov chain Monte-Carlo scheme. This algorithm governs the evolution of the parameters, along with the birth and death of segments. A simulated annealing scheme ensures the convergence to a clean and stable structure that corresponds to the vascular tree. The initial result is refined through a similar, but more accurate hierarchical polyline process focusing on border information to reconnect and extend the vascular network.

This approach is particularly appealing in the way prior knowledge on geometric primitives and on their interactions (connections and angles between segments, repulsive forces between proximate parallel branches) is elegantly expressed within an homogeneous, well-posed theoretical framework. This additional high-level knowledge, coupled with the flexibility of the stochastic process, is the key ingredient for both robust and accurate results. So far, this approach has only been applied on 2D data. Although there is no theoretical obstacle to its extension to 3D, we expect its already very high computational cost to increase significantly. This makes such a technique unfortunately prohibitive by today’s clinical standards.

4.6. Post-processing

Initial extraction results may be lacking in different aspects: surface information may be missing or inaccurate, the vessel’s topology may be incorrect, non-vessel regions may be included, vessel segments may be disconnected or missing. Hereafter, we briefly mention post-processing steps that were proposed to cope with such issues.

4.6.1. From centerlines to contours

Most centerline-based techniques (Section 4.4) extract the centerline without providing an accurate surface segmentation. The availability of the centerline structure greatly eases subsequent surface extraction through geometric and spatial constraints.

A first approach is to segment the vessel wall iteratively in cross-sectional planes. Among cross-section-dedicated techniques (Section 3.3), ray-casting schemes, 2D active contours (see also Fig. 28) and circular dynamic-programming schemes are good candidates for the accurate segmentation of cross-sectional contours.

From a collection of 2D contours, a 3D surface can be optimized using geometric models such as generalized cylinders (O’Donnell et al., 1994) and finite elements (O’Donnell et al., 1998) (see also Section 2.2). Another possibility is to extract or refine the surface directly in 3D: 3D parametric active surfaces are used in (Montagnat, 1999; Yim et al., 2001; Wong and Chung, 2006; Mille and Cohen, 2009), level-set evolutions in (van Bemmel et al., 2003), B-spline tensor surfaces in (Frangi et al., 1999a,b, 2000) (see Fig. 2). Another alternative, recently applied success-
fully to CTA coronary arteries, if the graph-cut-based approach of Schaap et al. (2009b).

4.6.2. From contours to centerlines
Recovering centerline information from a 3D binary segmentation is generally formulated as a skeletonization problem (Buhler et al., 2002; Suri et al., 2002b). Most skeletonization algorithms follow a homotopic thinning design (Malandain et al., 1993; Palagyi and Kuba, 1998), in which points are removed from the binary object in a particular order. Distance transforms are classically used (Metz et al., 2007; Pudney, 1998; Puig et al., 1997). Gradient flux maximization is a more recent alternative (Bouix et al., 2005; Siddiqi et al., 2002). Pruning strategies can be employed to remove spurious branches (Boskamp et al., 2004; Bouix et al., 2005; Wan et al., 2000; Wan et al., 2002) and centerline models can be applied for subsequent smoothing of the thinning result (see Section 2.2).

4.6.3. Spatial coherence and topology
Low-level pixel-wise techniques are particularly prone to issues such as holes, isolated false positives and noisy contours. Markov random fields regularization (Chung et al., 2004; Wong et al., 2004), maximum intensity projection analysis (Gan et al., 2004) or anisotropic smoothing (Yang et al., 2004) (see Fig. 13), help correct such problems. An alternative is the use of mathematical morphology: hole-filling is performed in (Wan et al., 2000, 2002), regulated morphology operators are used in (Agam et al., 2005; Wu et al., 2004).

4.6.4. Region competition and splitting
If an initial segmentation result includes non-vessel parts, region competition can be used to disconnect them from the actual vessel. We already mentioned competitive region-growing (Yi and Ra, 2003) and the hierarchical segmentation scheme from Holtzman-Gazit et al. (2006) in Sections 4.2 and 4.3, respectively. Another candidate technique is fuzzy connectedness and its various refinements (Udupa et al., 2002), successfully applied to artery-vein separation in (Lei et al., 2001).

4.6.5. Graph-based clean-up and reconnection
Graph techniques can also be used to clean-up cluttered segmentations, exploiting the intrinsic tree structure of vascular networks. A popular example is the use of minimum spanning tree algorithms, for which different valuation strategies have been proposed: Euclidian (Bullitt et al., 1999) or Mahalanobis distances (Jomier et al., 2005), energies accounting for intensity variations and path regularity (Florin et al., 2004). Various heuristics have been proposed to trim or reconnect centerline graphs (Lee et al., 2007; Szymczak et al., 2005; Yim et al., 2000, 2003) (see Fig. 29). In Agam et al. (2005); Wu et al. (2004), such heuristics are given for a graph of so-called fuzzy spheres, with the additional notion of radius similarities to value graph edges. Closely related, minimal path techniques can also be used to reconnect disconnected regions (Cohen and Deschamps, 2001; Deschamps and Cohen, 2002b).

4.7. General considerations on extraction schemes
The choice of an adequate extraction scheme depends notably on the degree of interactivity and type of output required, on the robustness, accuracy and computational efficiency needed. For instance, region-growing schemes are usually straightforward to implement, fast and minimally interactive. Their main issue is the design of robust inclusion and stopping criteria, which may be difficult in practice. Two-point centerline-based minimal path techniques have earned their popularity from their robustness, computational efficiency and control over boundary conditions. Counterparts are the increased need for manual interaction and, in classical implementations, the lack of surface segmentation. As emphasized throughout the document, good practice can be to combine extraction schemes. For instance, 1D centerlines can first be obtained from minimal paths and serve to initialize active contours for an accurate segmentation of the vessels’ surface (see Fig. 11).

5. Discussion
The medical interest in 3D vascular segmentation and the challenges it raises have motivated a tremendous amount of research.
work. Techniques from almost all image processing frameworks have been explored. The complexity of the methods proposed in the literature depends heavily on the difficulty of the application and on the levels of automation, accuracy and computational efficiency targeted. In practice, existing methods differ so greatly in their applicative frameworks and purposes that direct comparisons are often meaningless. Instead, the axes developed in this paper aim at clarifying the key design issues that led to the proposed techniques:

- What vascular-specific prior knowledge is expressed?
- How is it measured on the data? How reliable are these measurements?
- How are these measurements controlled, optimized, corrected?

As was illustrated in Table 2 in the early part of this paper, vascular segmentation methods are built upon a large variety of models, features and extraction schemes. By decomposing a whole method into elementary components, one gains insight on its core properties, strengths and limitations.

The use of a priori knowledge on the geometry and appearance of the target vessels is often crucial to the robustness and accuracy of a segmentation method. One can remark that model-based information is often embedded implicitly in the features and extraction schemes. Such concealed aspects must be taken into account to fully comprehend the practical behavior of a given segmentation framework.

To cope with most complex angiographic applications, advanced techniques often rely on combinations of several models and features, even mixing several extraction schemes. A way to improve existing algorithms in terms of both performance and automation can thus be the design of new sequential combinations: pre-processing can be employed to improve the quality of the image data, pre-segmentation to automatize the initialization of dedicated extraction schemes, rough delineation to constrain further refinements. More and more works in the literature consist in elaborating chains of existing components. Choosing the right component for each step of the process depends on computational constraints and, more importantly, on the prerequisites of the subsequent blocks. For instance, it may be particularly difficult to devise automatic initialization for an extraction algorithm with highly constrained input requirements.

Individual components can also be combined in a hierarchical fashion, with different levels corresponding to different abstractions. For instance, the method of Lorenz et al. (2003) is organized in three levels: voxel, segment and vascular tree. At the voxel level, only low-level appearance hypotheses are exploited. Assumptions of local tubularity are expressed naturally at the segment level, while the tree level controls bifurcations and centerline regularity. Another recent example can be found in (Socher et al., 2008), using a hierarchical marginal space paradigm trained by probabilistic boosting trees. Successive marginal spaces correspond respectively to the vessel positions, widths and lengths. Such layered designs clarify the inter-dependencies between the different components by making underlying abstractions explicit. We emphasize that the understanding of these dependencies, which are embedded implicitly in most techniques, is particularly important as they help identify potential performance bottlenecks. Ameliorations of inter-dependent designs can be obtained by improving or replacing individual components or by changing how critical pieces of information are extracted. For instance, 2D cross-sectional features (Section 3.3) require a test orientation as input. The local direction can be estimated from the data, thanks to other local features (risks of unreliability), by an exhaustive evaluation over a set of possible directions (potentially costly) or can be controlled at a higher level, e.g. predicted by a tracking extraction scheme.

Besides combinations of techniques, an interesting trend is the development of adaptive extraction schemes, underlining complex dynamics between models and features. Instead of relying on fixed formulations (e.g. a fixed threshold) to extract the entire vascular network, adaptive techniques consider each vessel as an individual target for which models, features and extraction parameters are automatically and dynamically adapted. In the recent literature, such methods have shown particular success at lowering both under- and over-segmentation issues. Tracking methods, in a broad sense, are particularly suited for adaptive schemes. Because of their iterative nature, these algorithms can easily exploit information acquired during the segmentation process. Adaptive schemes are often implemented as heuristic feedback loops controlling the adjustment of local parameters, e.g. dynamic thresholds. Recursive Bayesian approaches such as particle filters offer well-defined theoretical frameworks to guide local model estimations given image data while preserving the global coherence of the extracted result.

Aiming for a generic, flawless segmentation framework is probably illusory. The adequacy of a given method is of course highly dependent upon one’s practical requirements. The same solution might be overkill for simple applications and be seriously lacking for more difficult ones. Additionally, tailoring an algorithm with respect to the specificities of the application is generally necessary to reach reasonable levels of practical efficiency. We believe however in the general principle of building efficient, sophisticated collaborative frameworks out of simple components. As algorithmic designs tend to become more intricate, a deep understanding of the properties of each core component remains essential.

Acknowledgement

The authors would like to thank the reviewers for their constructive and thoughtful comments.

References


