Pattern Recognition 44 (2011) 1514–1527

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Detection and matching of curvilinear structures

C. Lemaitre [a], M. Perdoch [b], A. Rahmoune [a], J. Matas [b], J. Miteran [a,*]

[a] Laboratoire d'électronique, d'informatique et de l'image, LE2I, Faculté Mirande, Université de Bourgogne, 21000 Dijon, France
[b] The Center for Machine Perception, Faculty of Electrical Engineering, Czech Technical University, Technicka 2, 166 27 Prague 6, Czech Republic

ARTICLE INFO

Article history:
Received 19 April 2010
Received in revised form
10 January 2011
Accepted 13 January 2011
Available online 19 January 2011

Keywords:
Curvilinear structures
Wiry objects
Descriptor
Detector
Segmentation
Matching

ABSTRACT

We propose an approach to curvilinear and wiry object detection and matching based on a new curvilinear region detector (CRD) and a shape context-like descriptor (COH).

Standard methods for local patch detection and description are not directly applicable to wiry objects and curvilinear structures, such as roads, railroads and rivers in satellite and aerial images, vessels and veins in medical images, cables, poles and fences in urban scenes, stems and tree branches in natural images, since they assume the object is compact, *i.e.* that most elliptical patches around features cover only the object. However, wiry objects often have no flat parts and most neighborhoods include both foreground and background.

The detection process is first evaluated in terms of segmentation quality of curvilinear regions. The repeatability of the detection is then assessed using the protocol introduced in Mikolajczyk et al. [1]. Experiments show that the CRD is at least as robust as to several image acquisition conditions changes (viewpoint, scale, illumination, compression, blur) as the commonly used affine-covariant detectors. The paper also introduces an image collection containing wiry objects and curvilinear structures (the W–CS dataset).

1. Introduction

In this paper, we propose a method for detection, description and matching of wiry objects and curvilinear structures (CS) such as roads, railroads and rivers in satellite and aerial images, vessels and veins in medical images, cables, poles and fences in urban scenes, stems and tree branches in natural images; see Fig. 1 for examples.

Methods based on matching of distinctive local patches (known as interest points, distinguished regions, affine-covariant regions, keypoints) have achieved impressive results in many computer vision applications, such as stereo matching [2,3], object or face recognition [4–7] and image retrieval [8,9]. However, the methods are not directly applicable to wiry objects and curvilinear structures since they assume the object is locally planar and compact, *i.e.* that most elliptical neighborhoods of features cover only the object. But wiry object often do not include flat parts and most neighborhoods include both the foreground and the background.

Our objective is to design a detector and a descriptor which will play the role of a local patch and its descriptor for wiry objects. The proposed detection process consists of two main steps: cross-section detection and region construction. The cross-section detection exploits a non-standard gradient operator that takes into account differences in texture properties inside and outside a given region. A preliminary version of this detector was described in [10].

The region construction (*i.e.* growth in the axial direction) exploits constraints which allow eliminating false positives in highly textured images. The detection procedure is carried out in the scale space. Since affine transformations induce a shift and translation along any direction, most of the detection process is affine-invariant.

Axial points of the curvilinear structures are used as a set of stable points. For each of these points, we propose a variant of the shape-context descriptor, capturing information about local scale and orientation of the CS.

As a second contribution, we introduce a collection of images with a range of different wiry objects and curvilinear structures. We call the collection the W–CS dataset.

The rest of the paper is structured as follows. Section 2 summarizes the state-of-the-art for curvilinear region detection and matching. Details about the W–CS dataset are presented in Section 3. The curvilinear region detector (CRD) is described in Section 4. In Section 5, the CRD performance is evaluated on the problem of segmentation of curvilinear structures. The descriptor is introduced in Section 6. Section 7 studies the repeatability of

* Corresponding author.
E-mail addresses: cedric.lemaitre@u-bourgogne.fr (C. Lemaitre), perdom1@cmp.felk.cvut.cz (M. Perdoch), adel.rahmoune@gmail.com (A. Rahmoune), matas@cmp.felk.cvut.cz (J. Matas), miteranj@u-bourgogne.fr (J. Miteran).

0031-3203/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2011.01.005

Fig. 1. Examples of wiry objects and curvilinear structures.

the CRD detector following the protocol introduced by Mikolajczyk in [1] which allows comparing the CRD with commonly used affine-covariant detectors. Results for two-view matching of curvilinear structures are presented in Section 8. Finally, Section 9 concludes the paper.

2. Related work

Several approaches to the detection of curvilinear structures have appeared in the literature [11–15]. Most have been developed for a specific application, *e.g.* [16,15,17,18] are intended for detection of roads, railroads or rivers in satellite or aerial images and [19] for retinal vessel segmentation.

A general method for (curvi)linear structure was introduced by Steger [13]. As a first step for the detection of line profile, a Gaussian filter and its first- and second-order derivatives are calculated. The Gaussian filter allows locating the center of the cross-section, the edges of the CS are determined when the extremal value of the first derivative is reached, simultaneously with a zero crossing of the second-order derivative. Then, the resulting points are linked together into lines. This method is general, but detects CS with a narrow range of widths.

Carmichael et al. [20] uses a set of hierarchically organized edge queries learned from edge points of the detected object. During the learning phase, the object is observed from almost all possible viewpoints and segmented from the background. Decision trees based on edge queries of increasing diameter are used to capture the shape of each edge point. This approach is able to detect and localize objects in various indoor settings with a reasonable false positive rate. The drawback is the need to observe almost all possible appearances of the object, which makes it inefficient for indexing or recognition in real-world scenarios.

Mikolajczyk et al. [21] detects edges at multiple scales which are then described using a SIFT-like descriptor [4]. Dheng et al. [22] used a detector called the principal curvature-based region detector (PCBR), combined with the SIFT descriptors for object recognition. The authors defined the curvilinear region as a set of pixels inside a closed curve, obtained by the watershed algorithm. This definition does not include the use of geometric constraints on the final segmented region and might produce numerous false positives in textured regions.

Description of curvilinear structures is an open problem. Most of currently popular descriptors, such as SIFT [23], PCA-SIFT [24], shape context [8], steerable filters [25], differential invariants [26], spin images [27], complex filters [28], moment invariants [29], make use of all points inside a compact region that contains a given feature (a circle, a rectangle or an ellipse). Belongie et al. [8] employ a shape-context in order to describe the shape of the whole object. However, such a descriptor is not local and relies on the quality of the segmentation process. Therefore, we propose a novel descriptor which is closely related to both the shape context [8] and the idea of geometric hashing of Chum and Matas [30] who suggested to compute a geometric hash by using those features in the neighborhood that have similar properties as the described feature, *e.g.* having similar scale or orientation.

3. The dataset

n this study, experiments are conducted on both images coming from standard databases and a database we created, called W–CS dataset.[1] The set of images comprises three parts. The first part is a subset of the Berkeley database which has been widely used for image segmentation evaluation [31]. The second part is the DRIVE database [32] of retina vessels, for which the ground truth (gold standard) was obtained by manual segmentation and supervised by an experienced ophthalmologist. This database allows to compare quantitatively the performance of several detectors. The third part is the W–CS dataset, introduced in the paper. It includes 30 images, six groups of five images depicting the same scene, for repeatability evaluation and 28 sets of images of wiry objects intended primarily for matching experiments.

The 30 image subset allows evaluating the repeatability of curvilinear detectors according to the protocol analogical to Mikolajczyk et al. [1], *i.e.* to evaluate the robustness against viewpoint change, scale change, blur change, JPEG compression and lighting changes. For the viewpoint change, the camera direction is varied approximately from 0° to 60°. The blur change is obtained by varying the focus of the camera, whereas the scale change is obtained by varying the zoom. The scale factor ranges from 1.3 to 3.1. The JPEG compression was performed using Irfanview. The compression ratio varies from 60% to 97%. And finally, the lighting change is produced by varying the shutter speed of the camera. Fig. 11 depicts some of the images.

The 28 sets are images of wiry objects such as cables, fences, sticks, chairs, branches, etc. acquired from different viewpoints. The background differs significantly between images of the same object. Examples are presented in Fig. 1.

4. The curvilinear region detection

4.1. An overview

An overview of the curvilinear region detector is presented in the block diagram of Fig. 2. First, an edge detector based on the Canny operator [33,34] is applied. Next, the set of linked edges is searched for pairs satisfying certain conditions which become candidate cross-sections. Constraints are introduced on the local geometry of the detected region such as the local curvature, the

[1] The W–CS dataset will be public and available on the website of the authors.

local width variation, etc. in order to remove false positives. All these steps are performed in the scale space.

Redundant regions that occur at multiple scales are filtered. Finally, we apply merging and bridging operations, by exploiting the geometric and appearance properties of the endpoints of the detected regions, *e.g.* the collinearity, the width, and the luminance.

4.2. Detection of candidate regions

4.2.1. Cross-section detection

A curvilinear structure $\mathcal{C}$ is a set of contiguous boundary points detected in one-dimensional sections along an edge in a given image, as shown in Fig. 3(b). First, a continuous sequence of edge or ridge points in the image, *i.e.* points with local extrema of the gradient magnitude, is determined. Then, for each point in this sequence, we search for an interval, along the gradient direction, that contains two opposite gradients and a homogeneous part. The interval defines the section. A one-dimensional cross-section is depicted in Fig. 3(a).

The one-dimensional sections form a curvilinear region if their width is consistent with its neighboring sections, *i.e.* the variation of the width along the region is below a given threshold. By convention, a CS is represented either by a set of pairs of points or by using a set of axis points A_i, a local width w_i and an orientation of the edge θ_i that is perpendicular to the axis of the region:

$$\mathcal{C}=\{(R_i,L_i),\ i=0,\ldots,l\},\quad \mathcal{A}=\{(A_i,w_i,\theta_i),\ i=0,\ldots,l\}. \tag{1}$$

The parameter l denotes the length of the structure or the number of pairs, $l\in\mathbb{N}$. $L_i=(x_L,y_L)^T$ and $R_i=(x_R,y_R)^T$ are defined as the left and the right boundary points in the image space, respectively. And $x,y\in\mathbb{N}$ are the point coordinates in the image coordinate system of center Ω. The quantities A_i, w_i and θ_i are defined below:

$$A_i=\left(\frac{x_R+x_L}{2},\frac{y_R+y_L}{2}\right)^T,\quad w_i=\|R-L\|,\quad \theta_i=\tan^{-1}\left|\frac{y_R-y_L}{x_R-x_L}\right|. \tag{2}$$

The CS can be seen as a polygon of all the boundary points R_i, L_i, whereas the CR is regarded as the union of pixels inside the regions B_i, defined by the quadrilaterals $\{R_i,R_{i+1},L_{i+1},L_i\}$.

In order to define a section as a one-dimensional signal, the image coordinate system is mapped onto an approximation of the Frenet coordinate system, with Ω' as the center of the analyzed curve, as it is shown in Fig. 3(a). The approximation used for the Frenet coordinate system is the gradient orientation for each edge pixel. Some more precise measurements have been evaluated, using several points on the edge, without significant improvement of the results. Thus, for a given pixel in the image, the image signal is sampled along the line following the gradient direction. The one-dimensional signal $f(n):\mathbb{N}\to\mathbb{R}$ in the local coordinate system Ω' is then defined on the interval $n\in[1,N]$ of N integer samples. The local optimization process consists of finding n_L and n_R, respectively, the abscissae of the points L and R by checking Fourier "gradients" of the signal. To this end, four small windows of length λ (typically $\lambda=4$, 8 or 16) are considered (*c.f.* Fig. 3(c)). They are defined below:

$$v_L=f(n+n_L-\lambda),\quad v_R=f(n+n_L+w),$$

$$v_{SL}=f(n+n_L),\quad v_{SR}=f(n+n_L+w-\lambda).$$

Then, the DFT transforms V_L, V_R, V_{SL} and V_{SR} of each of these windows are computed. The correction by Hamming window is performed in order to reduce the boundary effects. The Fourier "gradient", which is the distance between two signals in the Fourier space, is expressed below:

$$m(V_1(k),V_2(k))=\frac{\sqrt{\sum_{k=0}^{\lambda}(|V_1(k)|-|V_2(k)|)^2}}{m_{max}},\quad V(k)=\sum_{n=0}^{\lambda}v(n)e^{-2i\pi kn/\lambda}. \tag{3}$$

The term m_{max} is a normalization factor and it corresponds to the maximum value of the distance m along one section, whereas

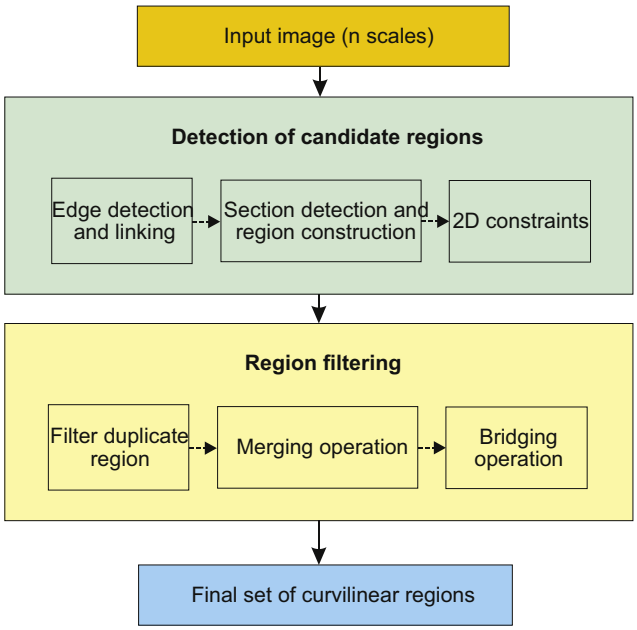

Fig. 2. The block diagram of the curvilinear region detector (CRD).

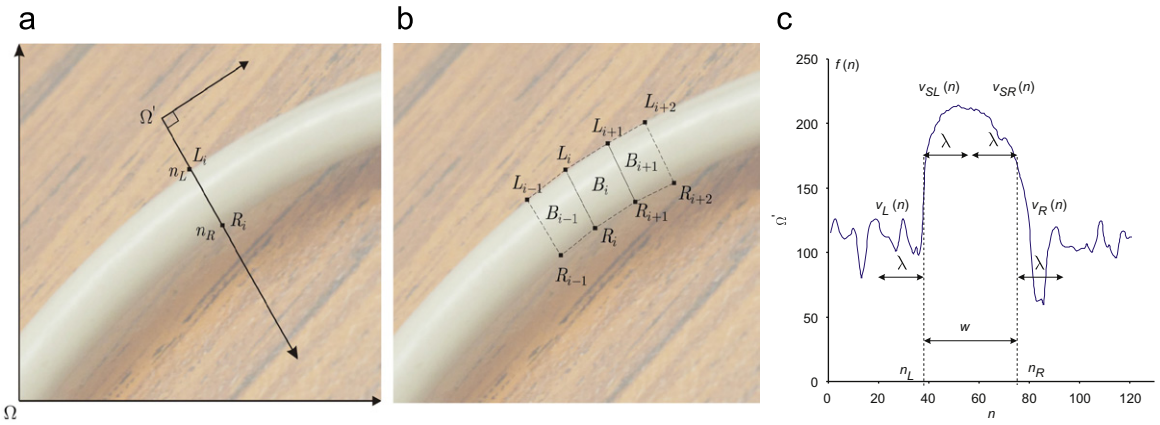

Fig. 3. A curvilinear region: (a) the coordinate system Ω' and one-dimensional section, (b) quadrilaterals of curvilinear structure and (c) an illustration of the one-dimensional cross-section detector.

$|V(k)|$ stand for the magnitude of the DFT. By using the Fourier "gradient" distance in combination with the usual local derivative of the intensity allows us to measure the distance between two textured patches, through exploiting the translation invariance property of the DFT [35].

The one-dimensional localization of the curvilinear section is posed as an optimization problem where the cost function, $Q(n_L, w)$, is defined as the product of four terms: the Fourier "gradient" on the left and the right side of the section, *i.e.* $P_L(n_L,w)$ and $P_R(n_L,w)$ respectively, the border similarity term $P_S(n_L, w)$ and the regularization term P_W. The quantities are given below:

$$P_L(n_L,w) = m(V_L(k),V_{SL}(k)),$$

$$P_R(n_L,w) = m(V_R(k),V_{SR}(k)),$$

$$P_S(n_L,w) = 1-m(V_L(k),V_R(k)).$$

The regularization term is a prior on the width of the curvilinear region w, which is modeled as a random variable according to the normal distribution, *i.e.* $P_W = N(w_p,\sigma_w)$. The mean w_p is equal to the width of the neighboring curvilinear section and the variance σ_w is related to the detector ($\sigma_w = 0.1$ in the next experiments).

The optimal parameters (n_L^*, w^*), which specify one-dimensional location of the section, are found by searching for the local maxima of the cost function $Q(n_L, w)$ according to the following equations:

$$Q(n_L,w) = P_L(n_L,w)P_R(n_L,w)P_S(n_L,w)P_W(n_L,w), \quad (4)$$

$$n_L^*,w^* = \underset{\substack{w \in \langle 0,w_{max} \rangle, \\ n_L \in \langle 0,n_{max}-2\lambda \rangle}}{\operatorname{argmax}} Q(n_L,w). \quad (5)$$

The parameter n_{max} stands for the length of the cross-section or the line cut; the parameter w_{max} is an upper bound on the width of the regions to be detected.

4.2.2. Curvilinear region construction

In order to form the curvilinear region, a curvilinear section is determined at every point along the edge line as described above. The sections are linked to form a curvilinear region. However, the resulting regions need to be filtered, so that only those having regular shapes are selected. The filtering process exploits geometric constraints.

4.2.3. The application of geometric constraints

Assuming that a given curvilinear region is well characterized by the width and the curvature, we introduced constraints on the continuity of both the width w_i and the curvature γ_i of the region. The curvature is estimated using the angle evaluated at the axis point A_i in the triangle defined by three consecutive axis points A_{i-1}, A_i, A_{i+1}. The constraints are defined in the following equation:

$$\frac{|w_i - w_{i+1}|}{\log(w_i)} < \varepsilon_1, \quad |\gamma_i - \gamma_{i+1}| < \varepsilon_2, \quad (6)$$

where ε_1 and ε_2 are the parameters of the detector (the thresholds were empirically determined and fixed to $\varepsilon_1 = 2$ and $\varepsilon_2 = 0.03$).

4.3. Multi-scale detection

All the region detection steps—edge detection, edge linking, cross-section detection, region construction and the application of two-dimensional constraints, are applied in the scale space. At each scale level k, the input image, $I(x,y)$, is smoothed with a two-dimensional Gaussian filter of variance σ_k to produce the image $I_{\sigma_k}(x,y)$, with $k = 1, \ldots, k_{max}$ (typically $k_{max}=3$). Then, the region detection steps are applied to each smoothed image, $I_{\sigma_k}(x,y)$, outputting curvilinear regions $\mathcal{C}_k$. The final set of regions is $\Phi = \bigcup \mathcal{C}_k$. The scale space approach allows detection of regions for a large range of widths (typically from $w=2$ pixels up to $w=100$ pixels in 1024×1024 images). This is also more computationally efficient than a single scale analysis with a high value of w_{max}. The parameters n_{max} and w_{max}, introduced in Eq. (5), are a linear function of the variance σ_k, and their initial value were usually fixed from 12 to 25 in the next experiments.

4.4. The merging and bridging operations

The detected regions are further analyzed for similarities, *e.g.*, in terms of neighborhood, collinearity, width, pixel values, etc. in order to decide whether to connect them.

The merging operation is defined as the construction of a new region C from two different regions A and B, with compatible properties on either of their extremities. The term "extremity" refers to a set of axis points, either at the head or at the tail of a given region. The extremities are denoted as A_1 and B_1, respectively. So, when an overlap is found between extremities, a multi-valued distance, $D_a = (d_1, d_2, d_3)$, is calculated. The term d_1 denotes the distance between the centers of gravity of the axis points of extremities A_1 and B_1, belonging to the overlapping area. The term d_2 denotes the relative difference between the average widths of A_1 and B_1. The last term d_3 denotes the difference between the orientations of A_1 and B_1. d_1 and d_2 are normalized regarding the average width of extremities.

If the distance D_a is small enough, *i.e.* $d_i \le \alpha_i$ with $i=1,\ldots,3$, then the encompassing extremity is preserved only in order to form the region C, in addition to the non-overlapping parts of both regions A and B. Even if there is some variation of the curvature in the overlapping area, d_1 is low if the two regions are "co-curvilinear", not necessarily "collinear", and d_1 is higher in other cases. Four cases are depicted in Fig. 4. This constraint

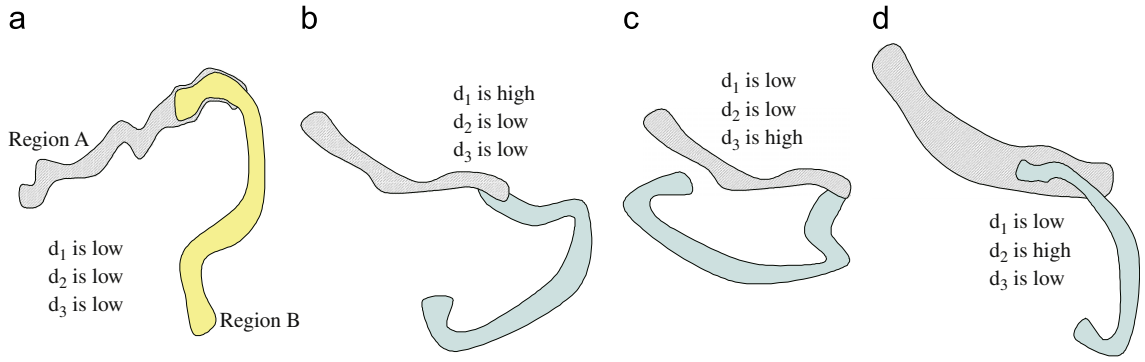

Fig. 4. An illustration of the merging operation: (a) the merging is possible, since A and B are "co-curvilinear", (b–d) the merging is not performed.

allows to remove some noise in complex background such as grass. In the retina context, the d_1 constraint could be released.

The thresholds, $(\alpha_1,\alpha_2,\alpha_3)$, are determined empirically and they define the merging model. We fixed $\alpha_1=0.4$, $\alpha_2=2$, $\alpha_3=45^\circ$ for the next experiments.

The bridging operation is defined as the construction of a region C from two distinct regions A and B, such that the geometric and appearance properties of either of their extremities are similar enough. The idea is similar to the method developed by Deschenes in [36] or by Steger in [13] for line junction detection, but we use a different definition of similarity, including properties such as width, collinearity and local luminance.

In the case that two disjoint extremities, A_1 and B_1, are found to lie close to each other, a global score function $F_a=f_1\times f_2\times f_3\times f_4$, is defined in order to find the most suitable extremities for bridging. The terms f_1 and f_2 are inversely proportional to the distances of the centers of gravity of both extremities, and their corresponding widths, respectively.

The quantity f_3 is defined in terms of the collinearity of both extremities. And finally, the term f_4 is related to the difference between the average luminance in the extremities A_1 and B_1. If each individual score is acceptable, *i.e.* $f_i\geq\beta_i$ with $i=1,\ldots,4$, and the global score F_a is local maximum over a search window, then, the two extremities A_1 and B_1 are extrapolated in order to form the bridged region C. It should be mentioned that the thresholds, β_i with $i=1,\ldots,4$, are determined empirically and they can be adjusted in order to deal with a range of bridging models ($\beta_1=4$, $\beta_2=4$, $\beta_3=15^\circ$, $\beta_4=3$ in the next experiments). Fig. 5 shows a typical example of the bridging operation, when applied on a real image.

4.5. Examples of region detection

Fig. 6 depicts some example results obtained by applying the curvilinear region detector. Identical parameters were set in all the cases (the value of key parameters is provided in Table 1). Most curvilinear regions of the objects are detected by the CRD despite the presence of complex or highly textured background. The performance of the detector is evaluated in detail in the next section.

5. Evaluation of the CRD detection

5.1. Qualitative evaluation of the detection

First, we compare the results of the CRD and the method of Steger [13], which is a general (unsupervised) method for detection of linear structure and its executable is available. The CRD parameters were set to optimize segmentation results on a subset of Berkeley database images. Fig. 7 shows detections produced by

Table 1
Key parameters of the CRD.

Step of detection	Parameters value
Scale	$w_{max}=12$, $k_{max}=3$
Geometric constraints	$\varepsilon_1=2$, $\varepsilon_2=0.03$
Merging	$\alpha_1=0.4$, $\alpha_2=2$, $\alpha_3=45^\circ$
Bridging	$\beta_1=4$, $\beta_2=4$, $\beta_3=15^\circ$, $\beta_4=3$

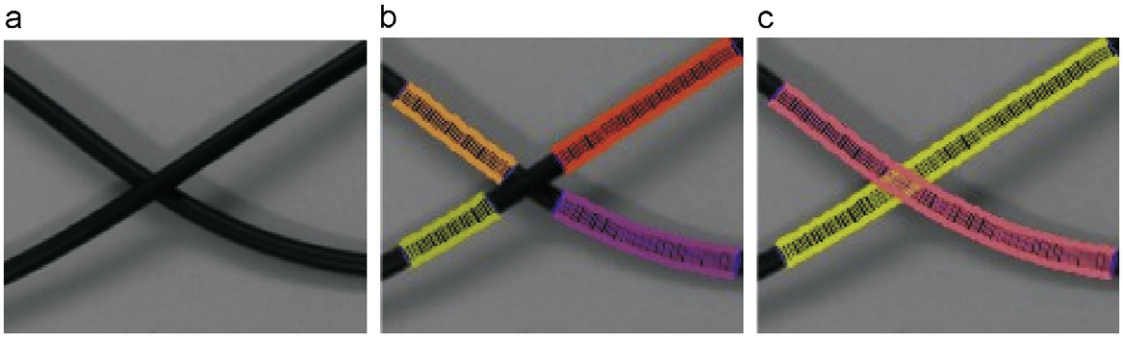

Fig. 5. An illustration of the bridging operation: (a) the original image showing the intersection of cables, (b) the detected regions without bridging, and (c) the detected regions are bridged.

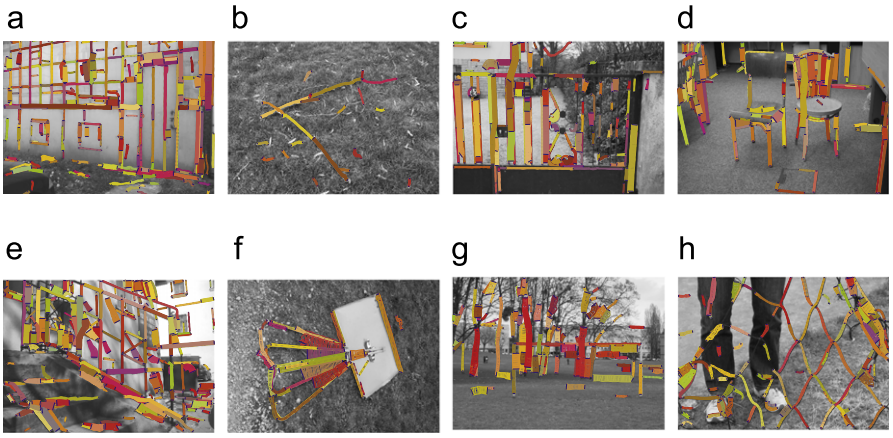

Fig. 6. Examples of the detected curvilinear regions.

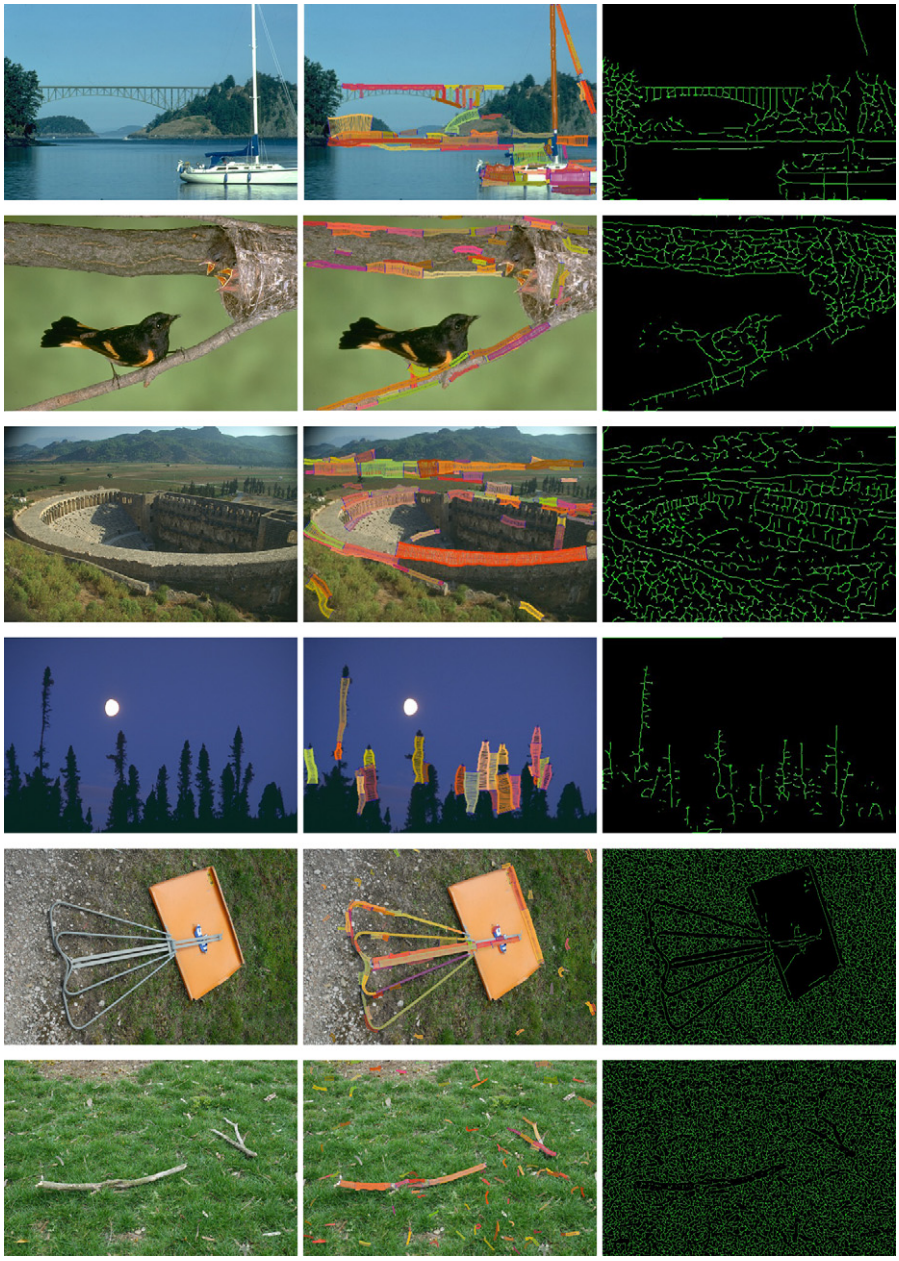

Fig. 7. Comparison of CRD and Steger's results. From left to right: original images, CRD results, Steger's results.

both methods. From left to right, the original image, the CRD result (edges and sections), and Steger's detector result (only the axis is available) are depicted.

Steger's detector produces many false positives in textured regions. This is particularly visible in the last two images, where numerous curvilinear regions are detected in the grass, as well as in the texture of the tree near the bird, or in the trees near the Roman amphitheater. Moreover, "thick" regions, *i.e.* regions with small length-to-width ratios, are not detected. The CRD is less sensitive to the background type and curvilinear objects are detected even on complex backgrounds, *e.g.* the sticks in Fig. 7 (bottom) and the stand of the sign in Fig. 7 (penultimate row).

5.2. Quantitative evaluation on retinal images

A potential application for the CRD is in medical image analysis, *e.g.* for the detection of blood vessels in retinal images.

Table 2
The precision of detectors for retina vessel segmentation.

Method	Precision P_s	Method	Precision P_s
Human	0.947	CRD	0.928
Staal	0.944	Jiang	0.921
Neimeijer	0.941	Martinez-Perez	0.9181
Zana	0.9377	Chaudhuri	0.87

To evaluate the CRD performance in this field, a thorough comparison against state-of-the-art detectors developed specifically for blood vessel segmentation, such as [37,38,2,39,32,40], was carried out. We used the reference database DRIVE [32], for which the ground truth was obtained through manual segmentation supervised by an experienced ophthalmologist. Note that the CRD is set to detect only dark curvilinear regions on light

background. This ensures the removal of some false positives between two blood vessels. Parameters of the CRD were tuned on a randomly chosen retinal image and then fixed for all experiments. Segmentation precision P_s is measured as the average number of well classified pixels, *i.e.* true positive pixels (TP) and true negative pixels (TN), which are present inside a given mask of size N_{mask}; $P_s = (TP + TN)/N_{mask}$. The processing is restricted to a circular mask encompassing all the blood vessels. The results of the precision for each detector are provided in Table 2.

The precision of the CRD (0.928) is very close to that of the best task-specific detector, which is Staal (0.944).

Fig. 8 shows two examples of the results of the retina vessel segmentation obtained by CRD, Chaudhuri, Martinez-Perez and Staal methods; manual annotation is also shown. The main differences in the results are due to the fact that the CRD does not detect vessels for which the contrast is very low. Note also that sometimes, the bridging step fails, mainly because of numerous cases of "Y" shape which is not modeled by our algorithm.

5.3. Quantitative evaluation in the general case

The test was carried out on images from the W–CS dataset, described in Section 3. "Ground truth" for these images was obtained by manual segmentation.

Detection results for a subset of W–CS dataset, namely for the branch, goal, fence, net, barrier, sign and stick images are depicted

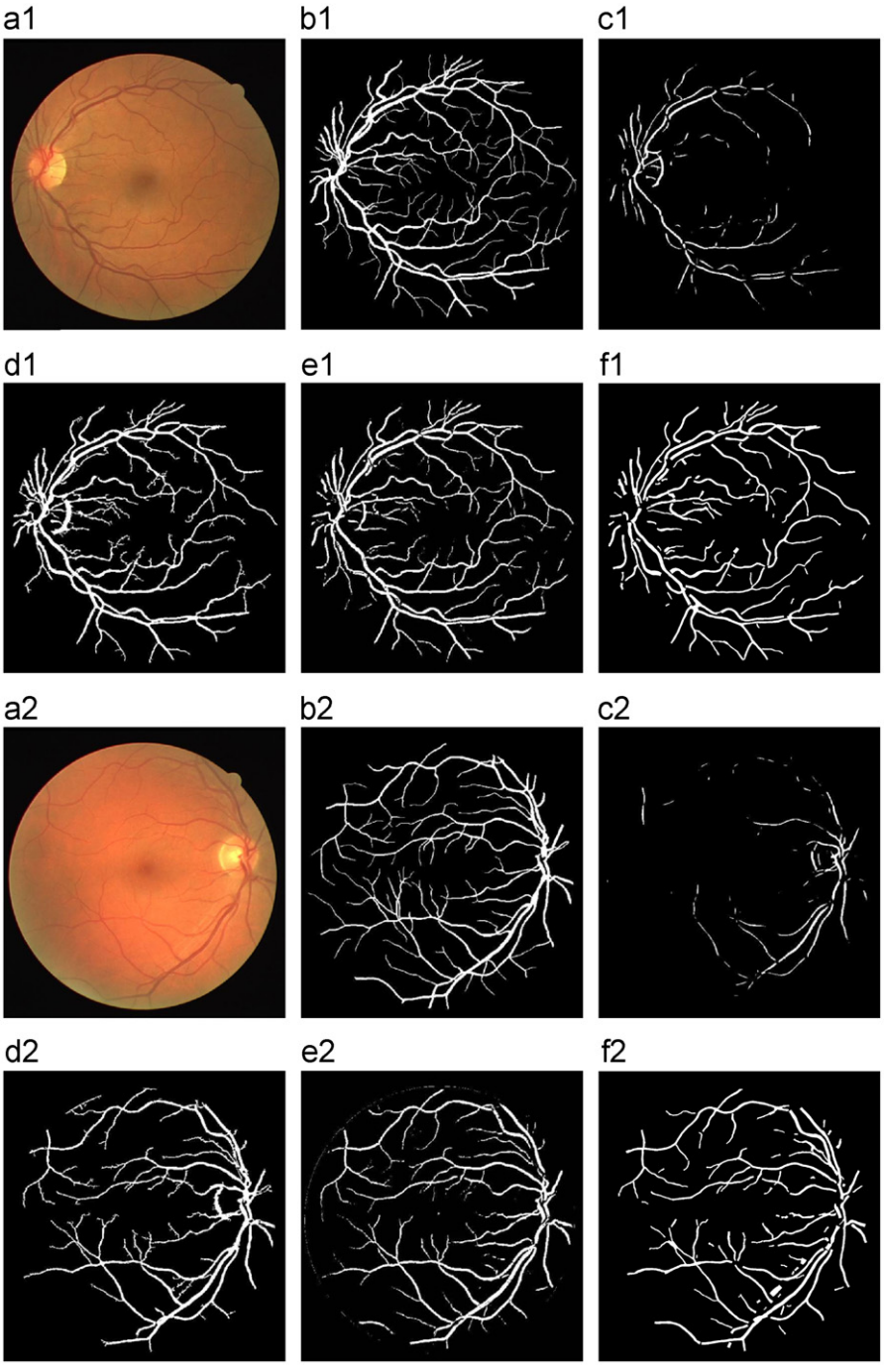

Fig. 8. Comparison of the segmentation results for the retina vessels: (a.1:2) the original images, (b.1:2) manual segmentation, (c.1:2) Chaudhuri method, (d.1:2) Martinez-Perez method, (e.1:2) Staal method, (f.1:2) CRD method.

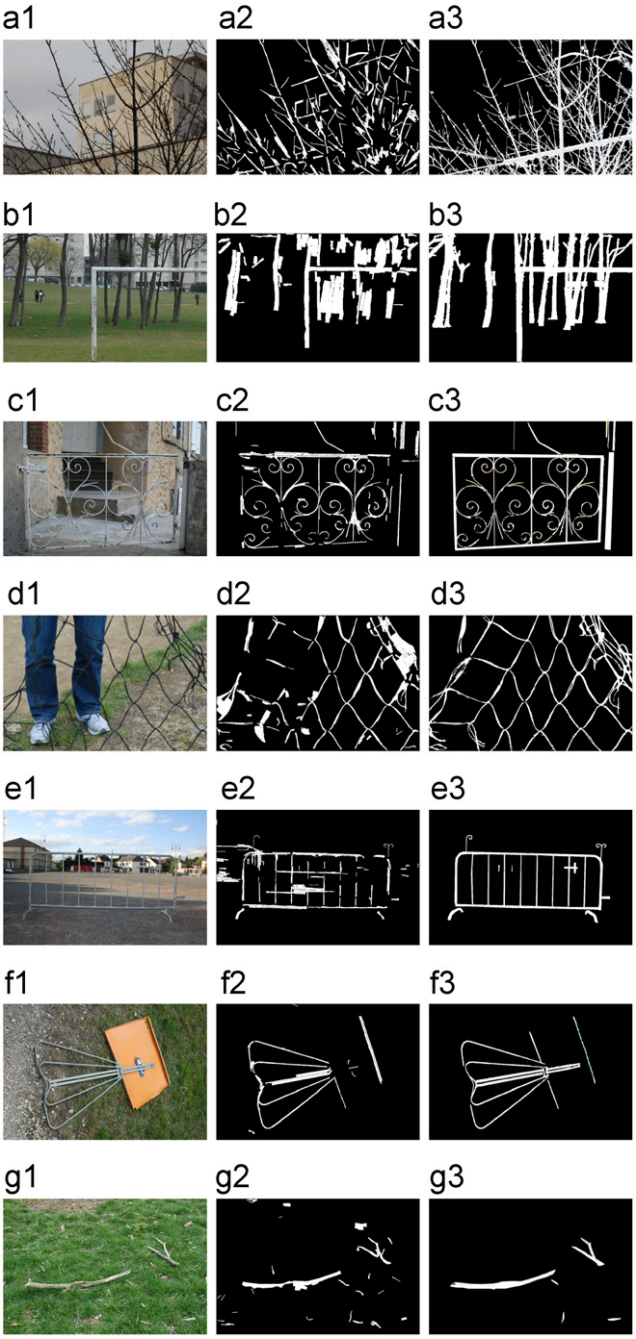

Fig. 9. Evaluation of the segmentation performances of the CRD. They are: the branch, the goal, the fence, the net, the barrier, the sign and the stick, downwards respectively. From left to right: (1) original images, (2) CRD results, (3) ground truth.

Table 3
The detection precision and the computation time of the CRD.

Image	Precision (P_s)	Time (s)
Branch	0.897	61
Goal	0.943	3
Fence	0.892	11
Net	0.912	19
Barrier	0.944	17
Sign	0.971	26
Stick	0.980	103

in Fig. 9. From left to right, the figure shows the original images, the result of the CRD, and the "ground truth" images.

Table 3 provides the segmentation precision (P_s) and the detection computation time. The precision numbers are in the range [89–100%], similar to those obtained in the specific context of retina vessel segmentation.

6. From detection to matching: the curvilinear region descriptor

Many points in a close vicinity of a curvilinear structure are unrelated to the curvilinear region itself. These points are a part of the background or another object. Thus, conventional region-based descriptors are not directly applicable for matching of curvilinear structures.

The proposed descriptor, which combines the idea of shape context [8] and geometric hashing [30], assumes that *curvilinear features with similar orientations and scale originated from the same object*. The idea can be extended to the similarity of intensity profile, color, texture or any other relevant measurement. The pre-segmentation according to properties facilitates matching on structured backgrounds, where standard descriptors that are sensitive to occlusions (patch-based descriptors) or assume segmented shape (shape-context) fail.

The curvilinear region shape descriptor is defined as follows. Let us recall the definition of a curvilinear structure

$$\mathcal{A} = \{(A_i, w_i, \theta_i), i = 0, \ldots, l\} \tag{7}$$

and denote the i-th element of a curvilinear structure $E_i = (A_i, w_i, \theta_i)$.

Let us assume that each element E_i is associated with a vector of measurement values $P_i = (p_{i1}, \ldots, p_{ik})$ that describes the properties of the element such as width and orientation, Fourier coefficients of the cross-section. Let us denote the similarity measure (*e.g.* Euclidean distance) between the properties of E_i and E_j as $d(P_i, P_j)$. Let $\mathcal{E}$ be the set of all curvilinear elements $E_j = (A_j, w_j, \theta_j)$, which are present in the neighborhood of E_i, within a radius $r_i = \xi w_i$, that possess properties P_j similar to the properties of P_i. Then the neighborhood of curvilinear element E_i is $\mathcal{E}_i = \{E_j | d(P_i, P_j) < \Delta, \|A_i, A_j\| < r_i\}$, where Δ is a maximum tolerated difference; the parameter ξ defines the size of the neighborhood.

The curvilinear descriptor is defined as a histogram of locations and orientations in the neighborhood $\mathcal{E}_i$. The locations are organized in a log-polar coordinate system as in GLOH descriptor [41]. However, instead of using gradient orientations, the oriented curvilinear elements $E_j \in \mathcal{E}_i$ are employed. Each curvilinear element, $E_j \in \mathcal{E}_i$, is added to the appropriate $(\theta_j - \theta_i)$ orientation bin in the log-polar coordinate system relative to E_i. The log-polar coordinate system is centered at the axis point A_i of E_i and the orientations are relative to the direction of θ_i. Additional robustness is gained by distributing values into neighboring bins both in orientation and position. Fig. 10 illustrates some regions, which are included in building the histogram.

In the experiments, the curvilinear orientation histogram descriptor (COH) employs eight orientation and three radii bins, each with histogram of four orientations, resulting in a 96-dimensional vector. The descriptor histograms are computed for curvilinear elements, sampled at a rate of 10 in order to alleviate the computational load. The size of the measurement region was set to $\xi = 8$. A typical length of curvilinear elements depends on the image size and the context, but is usually from $w_j = 2$ to 25 pixels at a given level of scale, and from $w_j = 2$ to 100 pixels using multi-scale detection. The typical number of curvilinear elements in a curvilinear structure is usually from $l = 20$ to 1000 for 1024 $\times$ 1024 images.

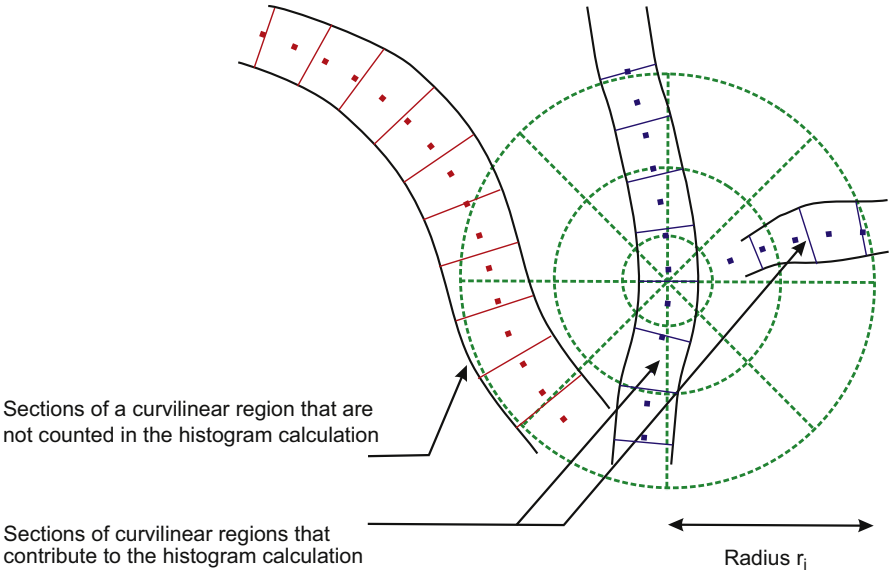

Fig. 10. An illustration of COH, the curvilinear region descriptor. The CS on the left is not included into COH as it has an incompatible (larger) width. The CS on the right has an acceptable width and it contributes to COH.

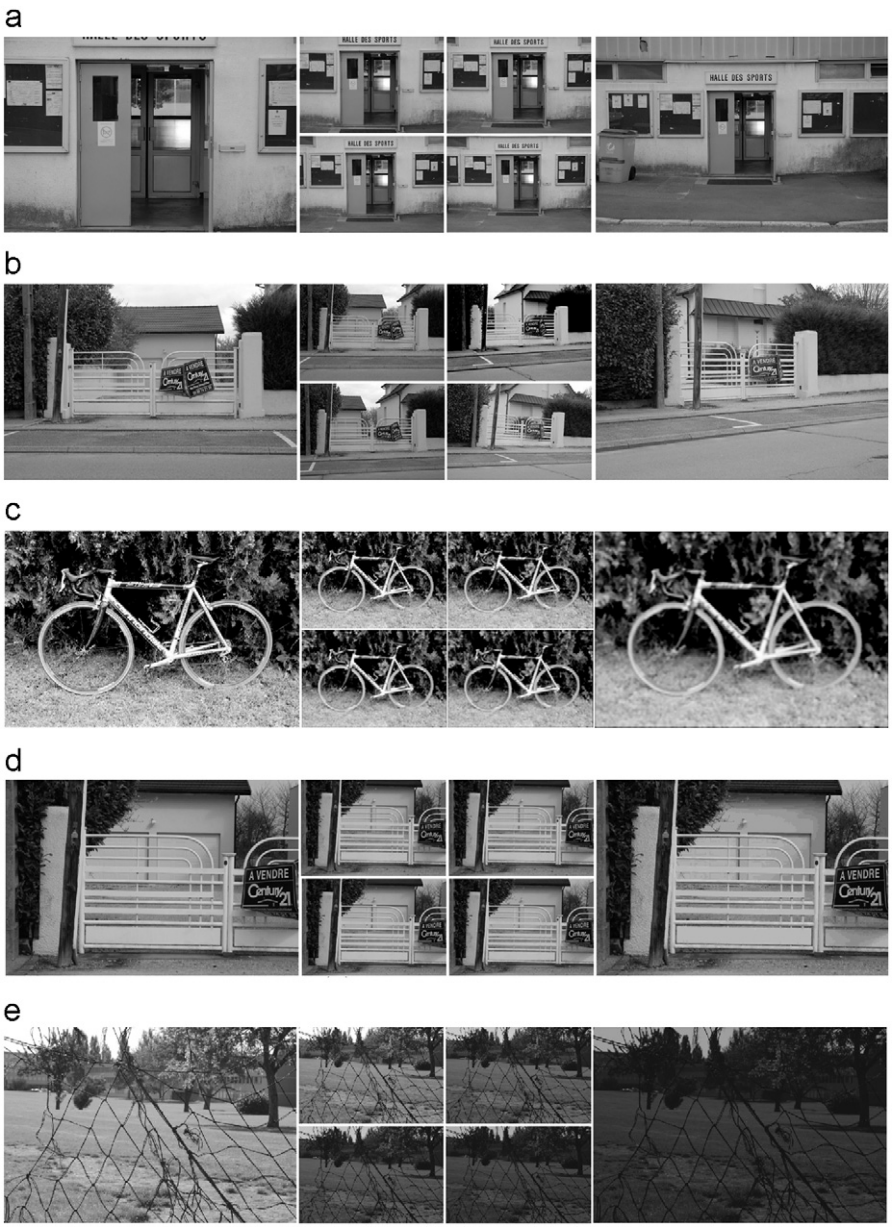

Fig. 11. Test images for repeatability evaluation under change of scale (a, first row) and viewpoint (b), with blur (c), JPEG compression (d), and illumination change (e, bottom).

7. Repeatability analysis

The aim of this analysis is to measure the robustness of the CRD detector to image transformations. By following the protocol of Mikolajczyk et al. [1] on W–CS images, we measure the repeatability of the CRD–COH with respect to: viewpoint change, scale change, blur change, JPEG compression and illumination change.

The CRD–COH repeatability is compared with established detection-description methods: MSER [2], IBR and EBR [3], Hessian-Affine [42] and Harris-Affine [43] detectors were used for this purpose, SIFT [4] for description.

The detected curvilinear structures were characterized by using the COH defined in Section 6. A set of matches is formed by finding all mutually closest pairs of features from the set of descriptors in the reference (first image) and in the distorted (viewpoint, scale, blur, JPEG compression and illumination change) image.

The test images are depicted in Fig. 11. For each experiment, a row of Fig. 11 depicts, from left to right, the reference image and the five distorted images. To evaluate repeatability, an overlap error introduced in the protocol [1] was used for standard detectors. With a known homography between the reference and the distorted image, detections from the distorted image are back-projected to the reference image and the overlap of elliptic regions is computed. The overlap error is defined as the ratio between the intersection and the union of two elliptic regions. For the curvilinear detector and descriptor, the overlap measure is not suitable because of the specific shape of the curvilinear structures. Therefore, a criteria that measures the distance between back-projections of curvilinear elements into reference image and the detections in reference image are used. Thus, the Figs. 12 and 13, showing the results of the repeatability performances for each method, allow mainly to compare the behavior of the detectors under different distortions.

In the matching part of the experiment, we have introduced a geometric consistency check. Each matching pair of curvilinear elements hypothesize a similarity transformation between images. The similarity transformation between each pair of matches is compared to the transformation of neighboring matches and inconsistent matching pairs are removed. We require each match to be consistent with at least 10 neighboring pairs. This extension to the matching part of the protocol is denoted as CRD/GC.

Scale change. To study the effects of a scale change on the CRD, a collection of images were obtained by varying the scale parameter in the following range: {1, 1.2, 1.35, 1.5, 1.95, 2.11}. These images are presented in the first row of Fig. 11.

The percentage of correct matches is high for both CRD approaches, ranging from 96% to 88% for the CRD with geometric consistency, *i.e.* CRD/GC, and from 77% to 63% for the plain CRD, *i.e.* without geometric consistency, see Fig. 12. Clearly, the CRD is at least as stable as other methods in terms of the percentage as well as in terms of the number of correct matches.

Viewpoint change. The set of images, which are employed for this test, are obtained by varying the viewpoint of the camera to take the following values: {0, 15, 25, 35, 45, 60} in degrees. The percentage of correct matches for the CRD/GC varies from 39% to 15%, whereas it decays from 27.2% to 8.8% for the plain CRD. The best percentage for a standard detector, methods, is that of the MSER where it varies from 1% to 7%. The decay of the CRD/CG is comparable to other methods in terms of number of correspondences.

Blur change. The test images were obtained by convolution with a Gaussian kernel of with the following values of σ : {0,2,3,4,5,6}. The decay of the CRD/CG is comparable to other methods in terms of the number of percentage of correct matches, and the CRD/CG outperforms clearly the CRD.

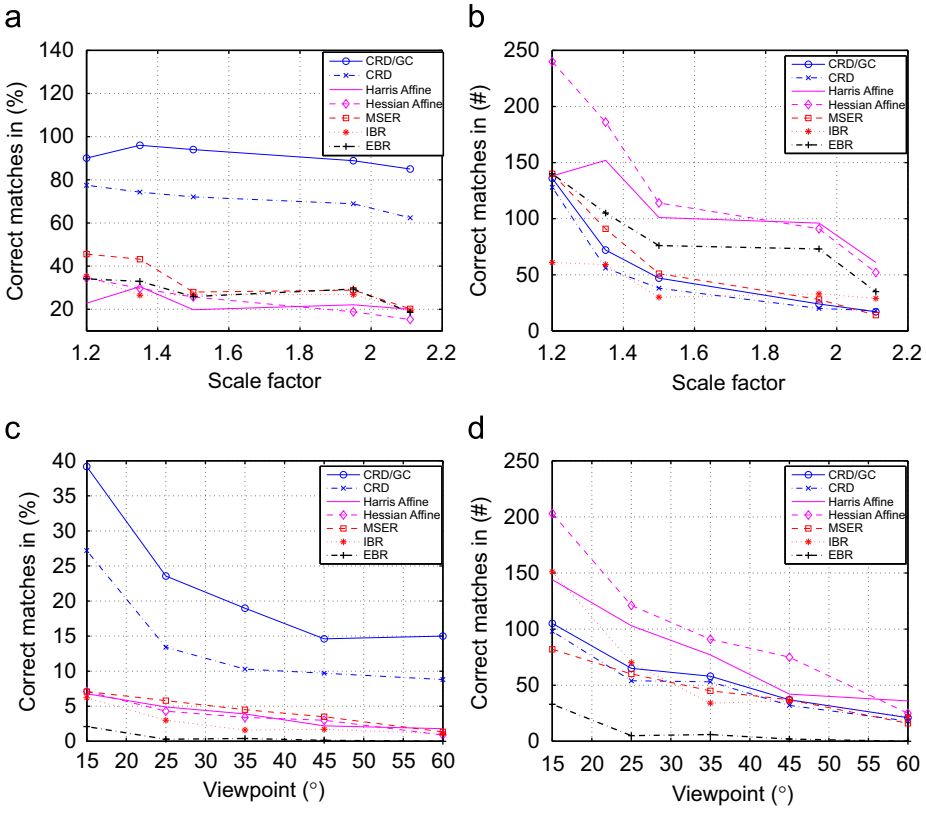

Fig. 12. Comparison of the repeatability performance for the CRD method, with respect to scale and viewpoint changes. The percentage (left) and the number (right) of correct matches.

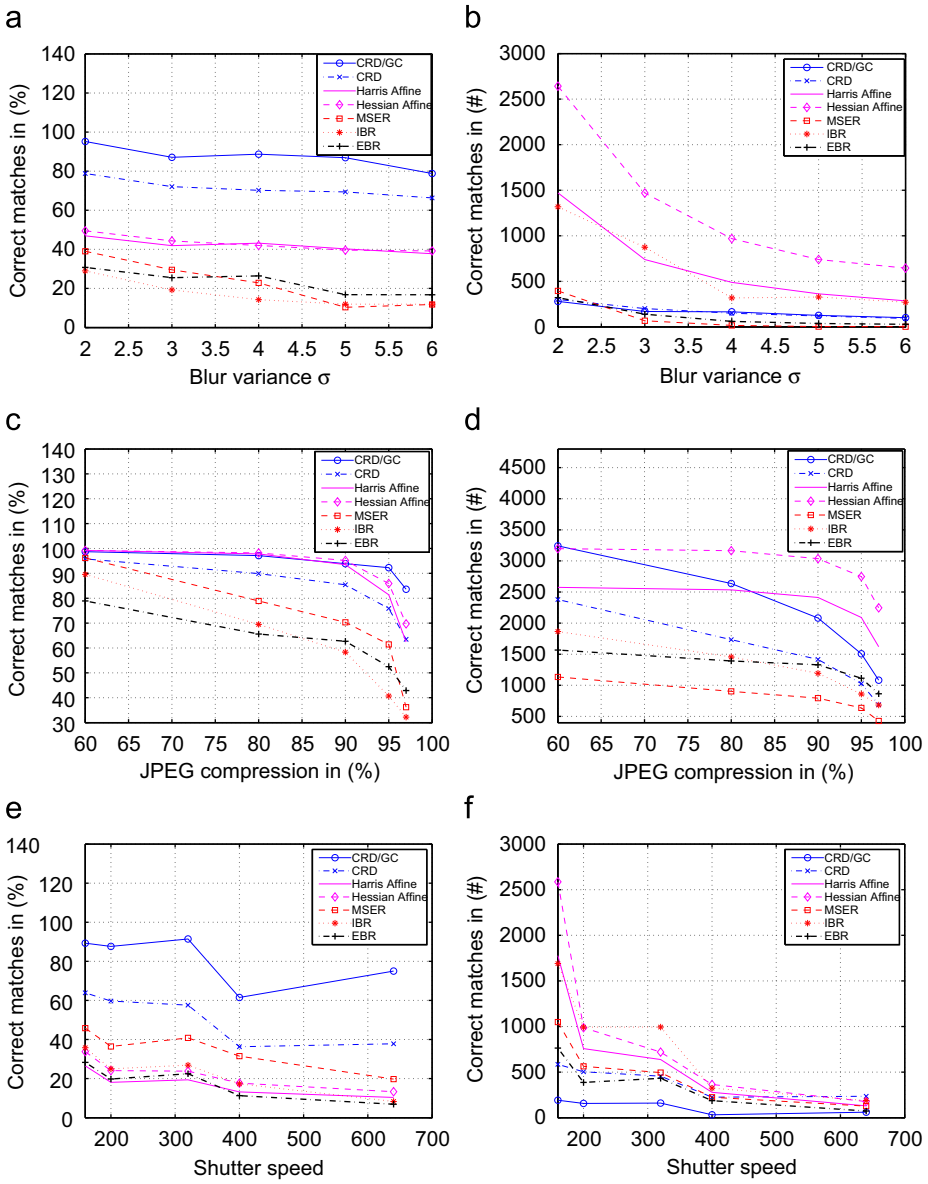

Fig. 13. Comparison of the repeatability performances for the CRD method, with respect to blur, JPEG compression and illuminations changes. The left figures represent the percentage of correct matches, whereas the right figures represent the number of correct matches.

JPEG compression change. The database of images was obtained by setting the JPEG compression ratio at different values. The image scene is chosen to contain a mixture of homogeneous regions as well as high-frequency parts. The percentage of correct matches of the CRD is very similar to the Hessian-Affine and the Harris-Affine methods. Note that with progressing JPEG compression, the overall number of correct matches decays significantly for the CRD as the quality of lower contrast edges decays due to JPEG compression artifacts. This is not a significant problem for the other region-based detectors.

Illumination change. The set of images was generated by changing the shutter speed of the camera. The exposure times were {1/100, 1/160, 1/200, 1/300, 1/400, 1/640} second. In terms of the percentage and number of correct matches, the CRD and CRD/CG are comparable to other detectors.

The number of correspondences, obtained using the CRD is usually lower than those obtained by using other methods. This is due to the fact that the number of curvilinear regions is naturally lower when compared to the number of features that are found in the other methods.

The most important result is that, except for the JPEG compression, the drop of the CRD related curves is generally smaller than for standard features.

8. Matching curvilinear structures

8.1. Matching

Pairs of still images from the W–CS dataset of natural scenes with wiry objects were employed in order to evaluate the matching performance of the CRD method. The image pairs are the chairs, the front of the house, the net, the fence, the cables and the sign. The pairs are characterized by slight changes in the viewpoint in scenes, which is sufficient to induce a significant background change.

The detected curvilinear structures were characterized by using the COH. A set of tentative correspondences was formed by finding all mutually closest pairs from the set of descriptors in the first and the second image, respectively.

Results are summarized in Table 4. The CRD–COH operator outperforms the DOG–SIFT method if the images are composed mainly of wiry objects and there is a significant background change between these images, *e.g.* for the cables, the net, the stairs and the front of the house. In such cases, the number of correct matches is much higher in the CRD–COH than in the DOG–SIFT approach. However, in the other cases, the number of correct matches in the CRD method is lower than for the DOG–SIFT. Examples of correct correspondences (the chairs, the front of the house, the net, the fence and the sign) are presented in Fig. 14.

8.2. The influence of the fourier gradient

We evaluated the effect of the window length λ of the Fourier gradient, used in the cross-section localization, on matching quality. The average results (for six pairs of images) obtained by varying λ from 1 to 16 are presented in Table 5.

For $\lambda=1$, the Fourier gradient is replaced by the standard first-order derivative. The optimum matching is obtained for $\lambda=8$. This experiment shows the superiority of the Fourier gradient compared to the standard first-order derivative.

Table 4
Comparison of the number and percentage of correct matches (inliers), obtained through the CRD–COH and the DOG–SIFT approaches.

	CRD–COH matches		DOG–SIFT matches	
	#	%	#	%
Fence	78	17	91	31
Cable a	413	69	370	89
Cable b	67	32	21	43
Chairs	76	29	136	79
Net	462	43	15	8
Stairs	32	4	1	1
Front of house	134	10	39	36

9. Conclusions

A method for the detection and description of wiry objects and curvilinear structures was presented. The contributions are: (1) a novel and general process for curvilinear region detection (CRD) and segmentation, (2) the introduction of the COH, a new shape-context like descriptor which allows matching of curvilinear structures and (3) the W–CS dataset.

We verified qualitatively and quantitatively that the CRD yields competitive segmentation results for different classes of images, such as retinal, urban scene and outdoor images. The repeatability of the detector combined with the descriptor was compared with commonly used affine-covariant detectors (*e.g.* MSER, Harris and Hessian-Affine). The detector of curvilinear regions is, with the exception of JPEG compression, less sensitive

Table 5
The effect of Fourier window length on the matching performance.

Window length λ	1	4	8	16
Correct matches (%)	16	26	35	30

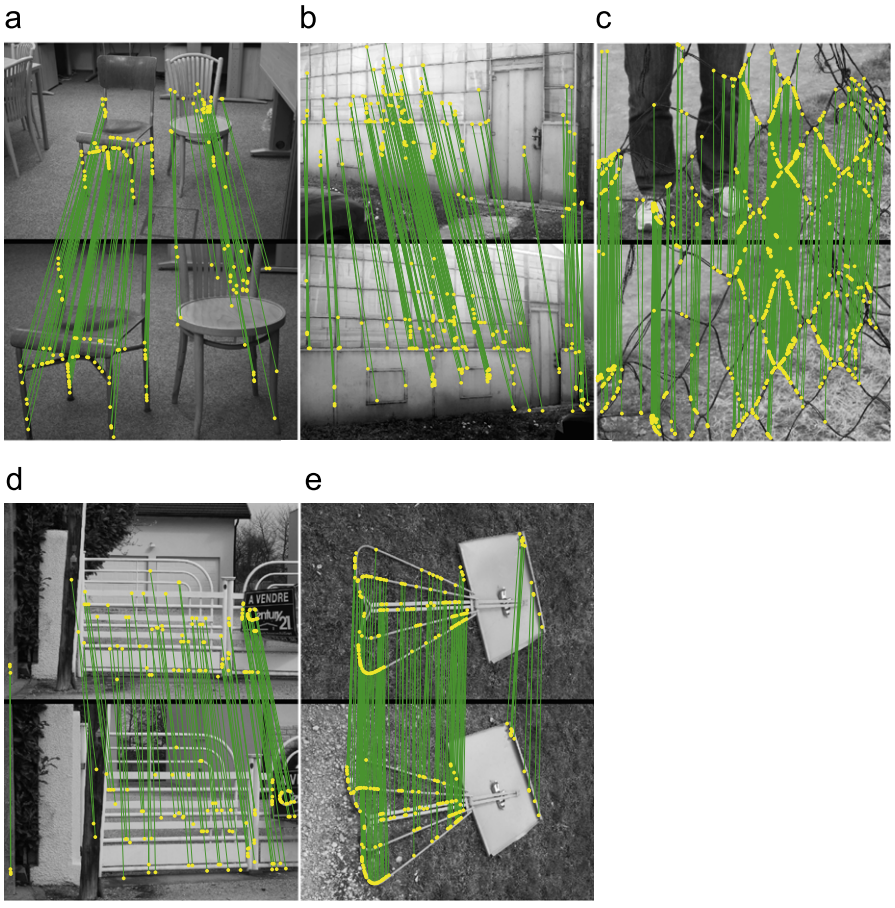

Fig. 14. Example of correct correspondences, which are found in scenes with viewpoint and background changes: (a) the chairs, (b) the front of the house, (c) the net, (d) the fence and (e) the sign.

to the image transformations (viewpoint, scale, illumination, compression, blur) than the standard detectors.

We demonstrated that the CRD–COH supports matching of wiry objects, even in difficult cases where the background is highly textured and when there are significant changes between the views. We note that the CRD–COH can be combined with standard features and descriptors to robustify image-to-image matching results, since the sets of patches are nearly disjoint.

Experiments also showed that the Fourier based gradient which takes into account differences in texture properties inside and outside of a given region outperforms the standard first-order derivative.

The processing time using a standard PC and C++ based implementation is around 20 s for 1024 $\times$ 1024 images. This is acceptable in many applications, but the speed does not support real-time use.

Acknowledgements

The authors were supported by Czech Science Foundation Project 102/07/1317 (JM), by EC project FP7-ICT-270138 Darwin (PM) and by Regional Council of Burgundy (CL, AR).

References

[1] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, L. Van Gool, A comparison of affine region detectors, International Journal of Computer Vision 65 (1–2) (2005) 43–72.
[2] J. Matas, O. Chum, M. Urban, T. Pajdla, Robust wide baseline stereo form maximally stable extremal regions, in: British Machine Vision Conference, 2002, pp. 384–393.
[3] T. Tuytelaars, L. Van Gool, Wide baseline stereo matching based on local, affinely invariant regions, in: British Machine Vision Conference, 2000, pp. 412–425.
[4] D.G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60 (2) (2004) 91–110.
[5] S. Obdrzalek, J. Matas, Object recognition using local affine frames on distinguished regions, in: British Machine Vision Conference, 2002, pp. 113–122.
[6] J. Sivic, A. Zisserman, Video Google: a text retrieval approach to object matching in videos, in: International Conference on Computer Vision, 2003, pp. 1470–1477.
[7] H.R. Kanan, K. Faez, Y. Gao, Face recognition using adaptively weighted patch PZM array from a single exemplar image per person, Pattern Recognition 41 (12) (2008) 3799–3812.
[8] S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape context, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (4) (2002) 509–522.
[9] O. Chum, J. Philbin, J. Sivic, M. Isard, A. Zisserman, Total recall: automatic query expansion with a generative feature model for object retrieval, in: International Conference on Computer Vision, 2007.
[10] C. Lemaitre, J. Miteran, J. Matas, Definition of a model-based detector of curvilinear regions, in: W.G. Kropatsch, M. Kampel, A. Hanbury (Eds.), Conference on Computer Analysis of Images and Pattern. Lecture Notes in Computer Science, vol. 4673, Springer, 2007, pp. 686–693.
[11] F. Deschênes, D. Ziou, Detection of line junctions in gray-level images, in: International Conference on Pattern Recognition, 2000, pp. 3762–3765.
[12] D. Ziou, Optimal line detector, in: International Conference on Pattern Recognition, 2000, pp. 3534–3537.
[13] C. Steger, An unbiased detector of curvilinear structures, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (2) (1998) 113–125.
[14] J.-H. Jang, K.-S. Hong, Detection of curvilinear structures and reconstruction of their regions in gray-scale images, Pattern Recognition 35 (4) (2002) 807–824.
[15] J. Mena, An automatic method for road extraction in rural and semi-urban areas starting from high resolution satellite imagery, Pattern Recognition Letters 26 (9) (2005) 1201–1220.
[16] A. Baumgartner, C. Steger, H. Mayer, W. Eckein, Multi resolution semantic and context for road extraction, in: Semantic Modeling for the Acquisition of Topographic from Images and Maps, Springer-Verlag, 1997, pp. 140–156.
[17] F.J.A. Lpez, J.L.G. Balboa, Generalization-oriented road line segmentation by means of an artificial neural network applied over a moving window, Pattern Recognition 41 (5) (2008) 1593–1609.
[18] N.S. Netanyahu, V. Philomin, A. Rosenfeld, A.J. Stromberg, Robust detection of straight and circular road segments in noisy aerial images, Pattern Recognition 30 (10) (1997) 1673–1686.
[19] T. Walter, J. Klein, P. Massin, F. Zana, Automatic segmentation and registration of retinal fluorescein angiographies—application to diabetic retinopathy, in: First International Workshop on Computed Assisted Fundus Image Analysis, 2000, pp. 1–6.
[20] O. Carmichael, Discriminative techniques for the recognition of complex-shaped objects, Ph.D. Thesis, Carnegie Mellon University, 2003.
[21] K. Mikolajczyk, A. Zisserman, C. Schmid, Shape recognition with edge-based features, in: British Machine Vision Conference, 2003, pp. 779–788.
[22] H. Dheng, W. Zhang, T. Dietterich, L. Shapiro, Principal curvature-based region detector for object recognition, in: IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8.
[23] D. Lowe, Object recognition from local scale-invariant features, in: International Conference on Computer Vision, 1999, pp. 1150–1157.
[24] Y. Ke, R. Sukthankar, Pca-sift: a more distinctive representation for local image descriptors, in: IEEE Conference on Computer Vision and Pattern Recognition, 2004, pp. 511–517.
[25] W. Freeman, E. Adelson, The design and use of steerable filters, IEEE Transaction on Pattern Analysis and Machine Intelligence 13 (9) (1991) 891–906.
[26] J. Koenderink, A. van Doorn, Representation of local geometry in the visual system, Biological Cybernetics 55 (1987) 367–375.
[27] S. Lazebnik, C. Schmid, J. Ponce, Sparse texture representation using affine-invariant neighborhoods, in: IEEE Conference on Computer Vision and Pattern Recognition, 2003, pp. 319–324.
[28] F. Schaffalitzky, A. Zisserman, Multi-view matching for unordered image sets, in: European Conference on Computer Vision, 2002, pp. 414–431.
[29] L. Van Gool, T. Moons, D. Ungureanu, Affine/photometric invariants for planar intensity patterns, in: European Conference on Computer Vision, 1996, pp. 642–651.
[30] O. Chum, J. Matas, Geometric hashing with local affine frames, IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, IEEE, 2006, pp. 879–884.
[31] D. Martin, C. Fowlkes, D. Tal, J. Malik, A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, in: International Conference on Computer Vision, vol. 2, 2001, pp. 416–423.
[32] J. Staal, M. Abramoff, M. Niemeijer, M. Viergever, B. van Ginneken, Ridge based vessel segmentation in color images of the retina, IEEE Transactions on Medical Imaging 23 (2005) 501–509.
[33] J. Canny, A computational approach to edge detection, IEEE Transactions on Pattern Analysis and Machine Intelligence 8 (6) (1986) 679–714.
[34] R. Deriche, Using canny's criteria to derive a recursively implemented optimal edge detector, International Journal of Computer Vision 1 (1987) 167–187.
[35] F. Smach, C. Lemaitre, J.P. Gauthier, J. Miteran, M. Atri, Generalized Fourier descriptors with applications to objects recognition in SVM context, Journal of Mathematical Imaging and Vision 30 (2008) 43–71.
[36] F. Deschênes, D. Ziou, Detection of line junctions and line terminations using curvilinear features, Pattern Recognition Letters 21 (6–7) (2000) 637–649.
[37] X. Jiang, D. Mojon, Adaptive local thresholding by verification-based multi-threshold probing with application to vessel detection in retinal images, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (1) (2003) 131–137.
[38] M. Martinez-Perez, A. Hughes, A. Stanton, S. Thom, A. Bharath, K. Parker, Scale-space analysis for the characterisation of retinal blood vessels, in: Medical Image Computing and Computer-Assisted Intervention — MICCAI'99, C. Taylor, A. Colchester (eds.), 1999, pp. 90–97.
[39] M. Niemeijer, J. Staal, B. van Ginneken, M. Loog, M.D. Abramoff, Comparative study of retinal vessel segmentation methods on a new publicly available database, in: I. Michae, J. Fitzpat rick, M. Sonka (Eds.), SPIE Medical Imaging, vol. 5370, SPIE, 2004, pp. 648–656.
[40] F. Zana, J. Klein, Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation, IEEE Transactions on Image Processing 10 (7) (2001) 1010–1019.
[41] K. Mikolajczyk, C. Schmid, A performance evaluation of local descriptors, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (10) (2005) 1615–1630.
[42] K. Mikolajczyk, C. Schmid, An affine invariant interest point detector, European Conference on Computer Vision 1 (2002) 128–142.
[43] K. Mikolajczyk, C. Schmid, Scale & affine invariant interest point detectors, International Journal of Computer Vision 60 (1) (2004) 63–86.

C. Lemaitre received a Ph.D. degree from the University of Burgundy in 2008 and after one year of post-doc in the University of Saint Etienne. He is currently working as a research fellow for the Fraunhofer Institute, Erlangen, Germany. During his Ph.D., he was involved in the development of a detector of curvilinear regions applied to matching.

M. Perdoch received the B.Sc. degree in software engineering from the Faculty of Electrical Engineering and Information Technology, Slovak University of Technology, Bratislava, Slovak Republic, in 2001 and the M.Sc. degree in computer science from the Faculty of Electrical Engineering, Czech Technical University, Prague, Czech Republic, in 2004. He is currently a Ph.D. student at the Center for Machine Perception. His current research interests include low level image processing, feature detection and description, object recognition and large scale object retrieval. He is a member of the IEEE.

A. Rahmoune received a Ph.D. degree from EPFL (Switzerland) in 2005. He worked on image/video representation and scalable coding using redundant dictionaries, and he works currently as a post-doc in the laboratory Le2i of University of Burgundy.

J. Matas received the M.Sc. degree (with honors) in cybernetics from the Czech Technical University, Prague, Czech Republic, in 1987 and the Ph.D. degree from the University of Surrey, Guildford, United Kingdom, in 1995. From 1991 to 1997, he was a research fellow in the Centre for Vision, Speech, and Signal Processing at the University of Surrey. In 1997, he joined the Center for Machine Perception at the Czech Technical University. Since 1997, he has held various positions at these two institutions. He has published more than 100 papers in refereed journals and conference proceedings. His publications have more than 800 citations in the Science Citation Index. He received the best paper prize at the British Machine Vision Conferences in 2002 and 2005 and at the Asian Conference on Computer Vision in 2007. He has served in various roles at major international conferences (*e.g.* IEEE International Conference on Computer Vision (ICCV), IEEE Conference on Computer Vision and Pattern Recognition (CVPR), International Conference on Pattern Recognition (ICPR), and Neural Information Processing Symposium (NIPS)), cochairing the European Conference on Computer Vision (ECCV) 2004 and CVPR 2007. He is an associate editor-in-chief of the IEEE Transactions on Pattern Analysis and Machine Intelligence and a member of the editorial board of the International Journal of Computer Vision. He is a member of the IEEE.

J. Miteran received a Ph.D. degree in 2004, and is a Professor at Laboratory Le2i of the University of Burgundy, France from 2006. His research interest is about real-time implementation of pattern recognition algorithms, and of classification methods, using reconfigurable computing (FPGA). He is on the program committee of international conferences such as HSPP, QCAV, SPIE Machine Vision Applications in Industrial Inspection, and ECCV Workshop Applications of Computer Vision 2004.