

Training of Templates for Object Recognition in Invertible Orientation Scores: Application to Optic Nerve Head Detection in Retinal Images

Erik Bekkers¹, Remco Duits¹, and Marco Loog²

¹ Department of Biomedical Engineering and Department of Mathematics and Computer Science, Eindhoven University of Technology, the Netherlands

² Pattern Recognition Laboratory, Delft University of Technology, the Netherlands

Abstract. A new template matching scheme for the detection of objects on the basis of orientations is proposed. The matching scheme is based on correlations in the domain $\mathbb{R}^2 \times S^1$ of complex valued invertible orientation scores. In invertible orientation scores, a comprehensive overview of how an image is decomposed into local orientations is obtained. The presented approach allows for the efficient detection of orientation patterns in an intuitive and direct way. Furthermore, an energy minimization approach is proposed for the construction of suitable templates. The method is applied to optic nerve head detection in retinal images and extensive testing is done using images from both public and private databases. The method correctly identifies the optic nerve head in 99.7% of 1737 images.

Keywords: template matching, multi-orientation, invertible orientation scores, optic nerve head, optic disk, retina

1 Introduction

We propose a new cross-correlation based template matching scheme for the detection of objects on the basis of local orientations. Template matching based on (normalized) cross correlation is a common approach to object recognition. The use of a similarity measure based on cross correlation is intuitive, easy to implement, and with the existence of optimization schemes for real-time processing [1, 2] a popular method to consider in computer vision tasks. However, the usual approach using pixel intensities as features for object recognition has its limitations, especially in applications where line-structures play an important role. In this case, template matching on the basis of geometrical information, e.g. local orientations, might be more appropriate (see e.g. [3]). We therefore generalize the concept of cross-correlations on position space \mathbb{R}^2 to the joint space $\mathbb{R}^2 \times S^1$ of positions and orientations ($\equiv SE(2)$, the Euclidean motion group). To this end, we represent an image $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ in the form of an *orientation score* $U_f : \mathbb{R}^2 \times S^1 \rightarrow \mathbb{C}$, i.e., a complex valued function on the extended domain $\mathbb{R}^2 \times S^1$. In an orientation score [4], we obtain a comprehensive overview

of how an image is decomposed into local orientations, see Fig. 1. We thus stay in the conventional and convenient framework of template matching via cross-correlation, however, the extension to orientation scores enables us to match patterns of orientation distributions, rather than pixel intensities.

For the construction of suitable templates we minimize an energy functional, where we pay attention to the following criteria:

1. The template should give a high response for inner products with positive object patches.
2. The template should ideally be perpendicular to negative object patches, i.e., the inner product with a negative object patch should be zero.
3. The template should be sufficiently smooth, as to prevent overfitting.

To enable 1 and 2 the energy functional contains a data-term. Here we make use of a representative training set of image patches, in which patches of the object of interest, as well as of objects not to be detected are included. To accommodate point 3, a Sobolev-type regularization-term is added to the energy functional. That is, regularization is done on the basis of gradients. In our extension to $SE(2)$ we make use of left-invariant gradients, i.e., a derivative frame that rotates with the orientation of the group elements $(\mathbf{x}, \theta) \in SE(2)$ in the orientation score. Consider to this end the $(\partial_\xi, \partial_\eta, \partial_\theta)$ -frame in the upper right figure of Fig. 1. The left-invariant gradients allow for (anisotropic) regularization in the direction of oriented structures.

The proposed generic template matching framework on $SE(2)$ is applied in the detection of the optic nerve head (ONH) in retinal images. Automated detection of the ONH is a challenging task and has therefore been the subject of many previous studies [5–11]. For a recent and extensive overview of ONH detection algorithms see [11]. Correct identification of the ONH is crucial in (automated) retinal image analyses, as the optic nerve head is either part of the analysis itself (classification of glaucoma [12]), or is used as a reference point in measurement protocols [13]. On conventional fundus (CF) images the ONH appears as a bright disk-like feature, but appears generally dark on images obtained by scanning laser ophthalmoscopy (SLO) cameras. Conventional ONH detection algorithms are designed for use with CF images, and are based on the analysis of pixel intensities [5, 6]. These approaches are fast, however, the performance typically decreases in the presence of pathologies. As an alternative, methods have been developed that include more contextual information and consider the typical pattern of blood vessels emerging from the optic nerve head [7–9]. These methods generally perform better than traditional methods; however, they often follow an elaborate processing pipeline, with high computational times as a consequence. Recently, methods have been proposed that are both fast and accurate, see [10, 11]. Our method is intuitive, easy to implement, fast and outperforms recent state-of-the-art methods on publicly available benchmark databases.

In this paper, we improve our recent work [14] on ONH detection by including: 1) training of templates; 2) regularization in $SE(2)$; and 3) a thorough investigation on the combination of complementary templates. The generic template matching framework is especially beneficial for the detection of objects

characterized by orientations/line structures, as is demonstrated in our application to optic nerve head detection in retinal images.

Structure of this Article. The remainder of this article is organized as follows: In Section 2 the reader is provided with the necessary prerequisites. The section starts with an explanation of orientation scores (Subsection 2.1), followed by normalized cross-correlation and the concept of cross-correlation on orientation scores (Subsection 2.2). An optimization scheme for the construction of templates for cross-correlation based template matching is provided (Subsection 2.3). Section 3 describes our approach to optic nerve head detection in retinal images. In Section 4 the performance of the method is reported and discussed. General conclusions can be found in Section 5.

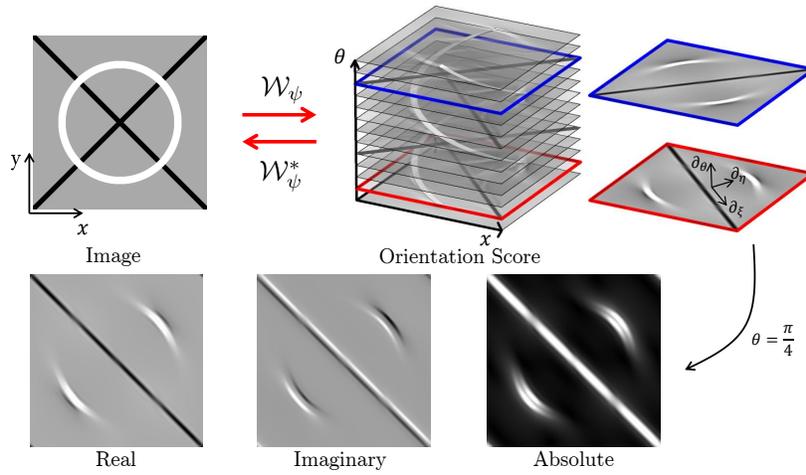


Fig. 1. Top row: Exemplary image and corresponding orientation score. Bottom row: Respectively the real part, imaginary part and modulus of a slice of the score at $\theta = \frac{\pi}{4}$.

2 Theory

2.1 Invertible Orientation Scores

An orientation score, constructed from image $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, is defined as a function $U_f : \mathbb{R}^2 \times S^1 \rightarrow \mathbb{C}$ and depends on two variables (\mathbf{x}, θ) , where $\mathbf{x} = (x, y) \in \mathbb{R}^2$ denotes position and $\theta \in [0, 2\pi]$ denotes the orientation variable. An orientation score U_f of image f can be constructed by means of correlation with some anisotropic wavelet ψ via

$$U_f(\mathbf{x}, \theta) = (\mathcal{W}_\psi f)(\mathbf{x}, \theta) = (\overline{\psi}_\theta \star f)(\mathbf{x}) = \int_{\mathbb{R}^2} \overline{\psi(\mathbf{R}_\theta^{-1}(\tilde{\mathbf{x}} - \mathbf{x}))} f(\tilde{\mathbf{x}}) d\tilde{\mathbf{x}}, \quad (1)$$

where $\psi \in \mathbb{L}_2(\mathbb{R}^2)$ is the correlation kernel, aligned with the x -axis, where \mathcal{W}_ψ denotes the transformation between image f and orientation score U_f , and \star denotes correlation. The overline denotes complex conjugation, $\psi_\theta(\mathbf{x}) = \psi(\mathbf{R}_\theta^{-1}\mathbf{x})$ and \mathbf{R}_θ is a counter clockwise rotation over angle θ . Note that $\tilde{\mathbf{x}} \in \mathbb{R}^2$ denotes a location in the image domain, whereas (\mathbf{x}, θ) denotes a location in the orientation score domain. The domain of an orientation score is essentially the classical Euclidean motion group $SE(2)$ of planar translations and rotations, equipped with product $g \cdot g' = (\mathbf{x}, \theta) \cdot (\mathbf{x}', \theta') = (\mathbf{R}_\theta \mathbf{x}' + \mathbf{x}, \theta + \theta')$.

In our work we choose cake wavelets [4] for ψ . Cake wavelets are designed to cover the entire Fourier domain, and have thereby the advantage over other oriented wavelets (s.a. Gabor wavelets) that they allow for a stable inverse transformation \mathcal{W}_ψ^* from the orientation score back to the image. As such, cake wavelets ensure that no data-evidence is lost during the transformation.

2.2 Template Matching via Normalized Cross Correlation

Normalized Cross Correlation in \mathbb{R}^2 . Let us consider a template and an image, $t, f : \mathbb{R}^2 \rightarrow \mathbb{R}$. We will denote translation by \mathbf{x} and rotation by θ of template t using the representation $(\mathcal{U}_g t)(\tilde{\mathbf{x}}) = t(\mathbf{R}_\theta^{-1}(\tilde{\mathbf{x}} - \mathbf{x}))$ and write $g = (\mathbf{x}, \theta) \in SE(2)$. The cross correlation coefficient as a function of translation and rotation of the template by g is then defined as follows:

$$c_{t,f}(g) = (\mathcal{U}_g t, f)_{\mathbb{L}_2(\mathbb{R}^2)} = \int_{\mathbb{R}^2} \overline{t(\mathbf{R}_\theta^{-1}(\tilde{\mathbf{x}} - \mathbf{x}))} f(\tilde{\mathbf{x}}) d\tilde{\mathbf{x}} = (\bar{t}_\theta \star f)(\mathbf{x}), \quad (2)$$

where $(\cdot, \cdot)_{\mathbb{L}_2(\mathbb{R}^2)}$ denotes the \mathbb{L}_2 inner product.

In order to make the correlation measure invariant to intensity scalings, both slots in the inner product can be normalized to zero mean and unit standard deviation. This is known as *normalized cross correlation*. To be able to normalize the image locally we make use of an additional mass function $m : \mathbb{R}^2 \rightarrow \mathbb{R}^+$ with $\int m(\tilde{\mathbf{x}}) d\tilde{\mathbf{x}} = 1$, which indicates the relevant region of the template, and define the $\mathbb{L}_2(\mathbb{R}^2)$ inner product using probability measure $m(\tilde{\mathbf{x}}) d\tilde{\mathbf{x}}$ as follows:

$$(t, f)_{\mathbb{L}_2(\mathbb{R}^2, m d\tilde{\mathbf{x}})} = \int_{\mathbb{R}^2} \overline{t(\tilde{\mathbf{x}})} f(\tilde{\mathbf{x}}) m(\tilde{\mathbf{x}}) d\tilde{\mathbf{x}}. \quad (3)$$

The normalized cross correlation coefficient $\hat{c}_{t,f}(g)$ is then defined as follows:

$$\hat{c}_{t,f}(g) = (\mathcal{U}_g \hat{t}, \hat{f}_g)_{\mathbb{L}_2(\mathbb{R}^2, \mathcal{U}_g m d\tilde{\mathbf{x}})}, \quad (4a)$$

$$\hat{t}(\tilde{\mathbf{x}}) = \frac{t(\tilde{\mathbf{x}}) - \langle t \rangle_m}{\|t - \langle t \rangle_m\|_{\mathbb{L}_2(\mathbb{R}^2, m d\tilde{\mathbf{x}})}}, \quad (4b)$$

$$\hat{f}_g(\tilde{\mathbf{x}}) = \frac{f(\tilde{\mathbf{x}}) - \langle f \rangle_{\mathcal{U}_g m}}{\|f - \langle f \rangle_{\mathcal{U}_g m}\|_{\mathbb{L}_2(\mathbb{R}^2, \mathcal{U}_g m d\tilde{\mathbf{x}})}}, \quad (4c)$$

with $\langle t \rangle_m = (1, t)_{\mathbb{L}_2(\mathbb{R}^2, m d\tilde{\mathbf{x}})}$ the local average with respect to the area covered by m , and with $\|\cdot\|_{\mathbb{L}_2(\mathbb{R}^2, m d\tilde{\mathbf{x}})} = \sqrt{(\cdot, \cdot)_{\mathbb{L}_2(\mathbb{R}^2, m d\tilde{\mathbf{x}})}}$.

Since the normalized image \hat{f}_g depends on g it needs to be calculated for every translation of the template, making this approach computationally expensive. Therefore, we will instead approximate (4c) by assuming that the local average is approximately constant in the area covered by m and that the mass is rotation invariant (i.e., $m(\mathbf{R}_\theta^{-1}\mathbf{x}) = m(\mathbf{x})$). That is, assuming $\langle f \rangle_{\mathcal{U}_{(\mathbf{x},\theta)}m}(\tilde{\mathbf{x}}) \approx \langle f \rangle_{\mathcal{U}_{(\tilde{\mathbf{x}},\theta)}m}(\tilde{\mathbf{x}}) = (m \star f)(\tilde{\mathbf{x}})$ for $\|\tilde{\mathbf{x}} - \mathbf{x}\|_{\mathbb{L}_2(\mathbb{R}^2)} < r$, with r the radius that determines the extent of m , we approximate (4c) as follows:

$$\hat{f}_g(\tilde{\mathbf{x}}) \approx \frac{f(\tilde{\mathbf{x}}) - (m \star f)(\tilde{\mathbf{x}})}{\sqrt{(m \star (f - (m \star f)))^2}(\tilde{\mathbf{x}})}. \quad (5)$$

Normalized Cross Correlation in $SE(2)$. Analogue to the \mathbb{R}^2 case, for two normalized orientation scores $\hat{T}, \hat{U}_f \in \mathbb{L}_2(SE(2))$ the normalized correlation is given by

$$\hat{C}_{T,U_f}(g) = \left(\mathcal{L}_g \hat{T}, \hat{U}_f \right)_{\mathbb{L}_2(SE(2)), \mathcal{L}_g M d\mathbf{x}d\theta}. \quad (6)$$

There we take the $SE(2)$ inner product with probability measure $M(\mathbf{x}, \theta) d\mathbf{x}d\theta$:

$$\left(\hat{T}, \hat{U}_f \right)_{\mathbb{L}_2(SE(2)), M d\mathbf{x}d\theta} = \int_{\mathbb{R}^2} \int_0^{2\pi} \overline{\hat{T}(\mathbf{x}, \theta)} \hat{U}_f(\mathbf{x}, \theta) M(\mathbf{x}, \theta) d\theta d\mathbf{x}, \quad (7)$$

and the shift-twist operator $(\mathcal{L}_g T)(\mathbf{x}, \theta) = T(\mathbf{R}_\alpha^{-1}(\mathbf{x} - \mathbf{b}), \theta - \alpha)$. Rotations by α followed by a translation \mathbf{b} via \mathcal{L}_g , with $g = (\mathbf{b}, \alpha)$, of orientation scores is done since $(\mathcal{W}_\psi \mathcal{U}_g f)(\mathbf{x}, \theta) = (\mathcal{L}_g \mathcal{W}_\psi f)(\mathbf{x}, \theta)$. Normalized template \hat{T} and orientation score \hat{U}_f are calculated in a similar fashion as described in Eq. (4b) and (5), where one can replace all inner products $(\cdot, \cdot)_{\mathbb{L}_2(\mathbb{R}^2, m d\tilde{\mathbf{x}})}$ by $(\cdot, \cdot)_{\mathbb{L}_2(SE(2)), M d\mathbf{x}d\theta}$ and where the correlation operator \star can be replaced by its $SE(2)$ equivalent:

$$(T \star_{SE(2)} U_f)(\mathbf{x}, \theta) = (\mathcal{L}_{(\mathbf{x},\theta)} T, U_f)_{\mathbb{L}_2(SE(2)), M d\mathbf{x}d\theta}. \quad (8)$$

Matching of Patterns of Orientation Distributions using $|U_f|$. Since both the orientation score transform (1) and template matching schemes, (4a) and (6), rely on a series of linear operators (correlations), it is possible to show that both Eq. (4a) and (6) produce the same results if the orientation score objects originate from their image equivalents. That is, there is no gain in performing template matching in $SE(2)$ if $U_f = \mathcal{W}_\psi f$ and $T = \mathcal{W}_\psi t$, since then $\operatorname{argmax}_{g \in SE(2)} \hat{c}_{t,f}(g) = \operatorname{argmax}_{g \in SE(2)} \hat{C}_{T,U_f}(g)$. However, in this work we find the ONH location $g_o = (\mathbf{x}_o, \theta_o) \in SE(2)$ via

$$g_o = \operatorname{argmax}_{g \in SE(2)} \left(\mathcal{L}_g \hat{T}, |\widehat{U}_f| \right)_{\mathbb{L}_2(SE(2)), M d\mathbf{x}d\theta}. \quad (9)$$

Here template matching in $SE(2)$ is done via the modulus of the orientation scores. This adaptation makes that the *appearance* (encoded in the phase) of

structures is not measured, it is rather the *presence* of structures that is being detected, consider to this end the bottom row of Fig. 1. We remove the image DC component before applying the orientation score transform. This guarantees a low response at locally constant regions where no orientation preference is expected. The absolute orientation score $|U_f(\mathbf{x}, \theta)|$ can then be regarded as a measure for finding an oriented structure at position \mathbf{x} and orientation θ . Note also that similar techniques for linear structure detection have been used before by Freeman et al. using (steerable) quadrature filter pairs [15].

2.3 Template Training

We describe a framework for the construction of suitable templates via the minimization of an energy functional. First, the energy functionals for both the \mathbb{R}^2 and the $SE(2)$ are described. Then, the templates will be represented in a B-spline basis. This allows for efficient and accurate optimization of the energy functionals. Finally, the minimizers corresponding to the energy functionals are presented in matrix-vector notation. A simple conjugate gradient approach can be used to solve for the B-spline coefficients.

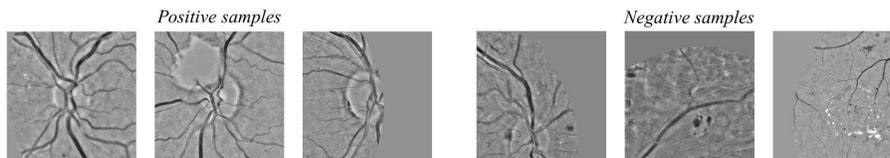


Fig. 2. Exemplary retinal image patches used for template training.

Energy Functional. For the optimization of 2D image templates t , the following energy functional is minimized:

$$E(t) = \underbrace{\sum_{i=1}^P \left(\left(t, \hat{f}_i \right)_{L_2(\mathbb{R}^2, m d\bar{\mathbf{x}})} - y_i \right)^2}_{\text{data term}} + \lambda \underbrace{\iint_{\mathbb{R}^2} \|\nabla t\|_{L_2(\mathbb{R}^2)}^2 dx dy}_{\text{regularization term}}, \quad (10)$$

where \hat{f}_i is one of P normalized patches representing either the optic nerve head, in which case the corresponding label $y_i = 1$, or a negative samples, in which case $y_i = 0$. Patches with label $y_i = 1$ will be referred to as positive patches, those with label $y_i = 0$ as negative patches. Examples of training patches are given in Fig. 2. The data-term of the functional trains the template to give a response of 1 if the inner-product is taken with a positive patch, and to give response 0 otherwise. The regularization term ensures that the template is smooth enough. I.e.,

sharp transitions in image intensities are punished using the squared gradient magnitude $\|\nabla t\|^2$. Parameter λ balances the data- and regularization-term.

Similar to Eq. (10), for the optimization of the orientation score template T the following functional is minimized:

$$\mathcal{E}(T) = \underbrace{\sum_{i=1}^P ((T, \hat{U}_{f_i})_{\mathbb{L}_2(SE(2), Mdx d\theta)} - y_i)^2}_{\text{data term}} + \lambda \underbrace{\iiint_{SE(2)} \|\nabla T\|_D^2 dx dy d\theta}_{\text{regularization term}}, \quad \text{with}$$

$$\iiint_{SE(2)} \|\nabla T\|_D^2 dx dy d\theta = \iiint_{SE(2)} D_{\xi\xi} \left| \frac{\partial T}{\partial \xi} \right|^2 + D_{\eta\eta} \left| \frac{\partial T}{\partial \eta} \right|^2 + D_{\theta\theta} \left| \frac{\partial T}{\partial \theta} \right|^2 dx dy d\theta, \quad (11)$$

and with the left-invariant gradient $\nabla T = \left(\frac{\partial T}{\partial \xi}, \frac{\partial T}{\partial \eta}, \frac{\partial T}{\partial \theta} \right)^T$ defined by

$$\partial_\xi := \cos \theta \partial_x + \sin \theta \partial_y, \quad \partial_\eta := -\sin \theta \partial_x + \cos \theta \partial_y, \quad \text{and } \partial_\theta. \quad (12)$$

Note that ∂_ξ gives the spatial derivative in the direction aligned with the orientation score kernel used at layer θ , recall Fig. 1. The parameters $D_{\xi\xi}$, $D_{\eta\eta}$ and $D_{\theta\theta}$ are used to balance the regularization in the three directions. Similar to this problem, first order Tikhonov-regularization on $SE(2)$ is related³, via temporal Laplace transforms, to left-invariant diffusions on the group $SE(2)$. In which case $D_{\xi\xi}$, $D_{\eta\eta}$ and $D_{\theta\theta}$ denote the diffusion constants in ξ , η and θ direction. Here we set $D_{\xi\xi} = 1$, $D_{\eta\eta} = 0$, and thereby we get Laplace transforms of hypo-elliptic diffusion processes [16, 17]. Parameter $D_{\theta\theta}$ can be used to tune between isotropic (large $D_{\theta\theta}$) and anisotropic (low $D_{\theta\theta}$) diffusion. See Fig. 3, where we have illustrated the Green's function of hypo-elliptic diffusion processes and the effect of regularization parameter $D_{\theta\theta}$ in the score domain. Note that anisotropic diffusion, via a low $D_{\theta\theta}$, is preferred as we want to maintain line structures in orientation scores.

B-Spline Basis. In order to efficiently minimize (10) and (11), the templates are described in a B-spline basis of direct products of n -th order B-splines B^n :

$$t(x, y) = \sum_{k=1}^{N_k} \sum_{l=1}^{N_l} c_{k,l} B^n \left(\frac{x}{s_k} - k \right) B^n \left(\frac{y}{s_l} - l \right), \quad (13a)$$

$$T(x, y, \theta) = \sum_{k=1}^{N_k} \sum_{l=1}^{N_l} \sum_{m=1}^{N_m} c_{k,l,m} B^n \left(\frac{x}{s_k} - k \right) B^n \left(\frac{y}{s_l} - l \right) B^n \left(\frac{\theta \bmod 2\pi}{s_m} - m \right), \quad (13b)$$

with $B^n(x) = \left(1_{[-\frac{1}{2}, \frac{1}{2}]} \ast^{(n)} 1_{[-\frac{1}{2}, \frac{1}{2}]} \right) (x)$ a n -th order B-splines obtained by n -fold convolution of the indicator function $1_{[-\frac{1}{2}, \frac{1}{2}]}$, and $c_{k,l}$ and $c_{k,l,m}$ the coefficients belonging to the shifted B-splines for \mathbb{R}^2 respectively $SE(2)$.

³ In which case $\|T - U_f\|_{\mathbb{L}_2(SE(2))}^2$ is used for the data term instead of the one in (11).

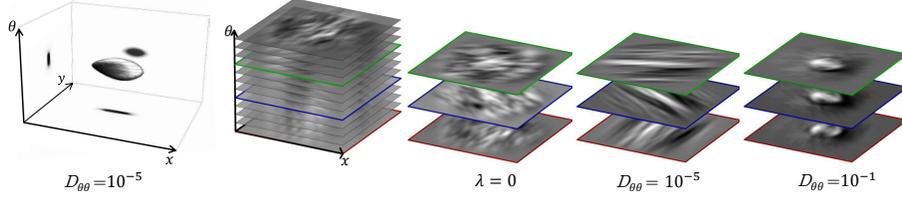


Fig. 3. The Green's function of hypo-elliptic diffusion with $D_{\theta\theta} = 10^{-5}$ and an orientation score template with different regularization settings. From left to right: no regularization ($\lambda = 0$), anisotropic ($D_{\theta\theta} = 10^{-5}$), and isotropic regularization ($D_{\theta\theta} = 10^{-1}$).

The Minimizer for the \mathbb{R}^2 Case in Matrix-Vector Notation. By substitution of (13a) in (10), the energy functional can be expressed in matrix-vector notation as follows:

$$E(t) = E_D(\mathbf{c}) := \|\mathbf{S}\mathbf{c} - \mathbf{y}\|^2 + \mathbf{c}^\dagger \mathbf{R}\mathbf{c}. \quad (14)$$

The corresponding minimizer is given by

$$(\mathbf{S}^\dagger \mathbf{S} - \lambda \mathbf{R})\mathbf{c} = \mathbf{S}^\dagger \mathbf{y}. \quad (15)$$

Here \mathbf{S} is a $[P \times N_k N_l]$ matrix given by

$$S = \{(s_{1,1}^i, s_{1,2}^i, \dots, s_{1,N_l}^i, s_{2,1}^i, s_{2,2}^i, \dots, s_{2,N_l}^i, \dots, \dots, s_{N_k,2}^i, \dots, s_{N_k,N_l}^i)\}_{i=1}^P, \quad (16)$$

$$s_{k,l} = (B_{s_k s_l}^n * (m \hat{f}_i))(k, l),$$

with $B_{s_k s_l}^n(x, y) = B^n\left(\frac{x}{s_k}\right) B^n\left(\frac{y}{s_l}\right)$, for all (x, y) on the discrete spatial grid on which the discrete input image $f_D : \{1, N_x\} \times \{1, N_y\} \rightarrow \mathbb{R}$ is defined. Here N_k and N_l denote the number of splines in resp. x and y direction, and $s_k = \frac{N_x}{N_k}$ and $s_l = \frac{N_y}{N_l}$ are the corresponding resolution parameters. The $[N_k N_l \times 1]$ column vector \mathbf{c} contains the B-spline coefficients, and the $[P \times 1]$ column vector \mathbf{y} contains the labels, stored in the following form

$$\mathbf{c} = (c_{1,1}, c_{1,2}, \dots, c_{1,N}, c_{2,1}, c_{2,2}, \dots, c_{2,N}, \dots, \dots, c_{M,2}, \dots, c_{M,N})^T$$

$$\mathbf{y} = (y_1, y_2, \dots, y_P)^T. \quad (17)$$

The $[N_k N_l \times N_k N_l]$ regularization matrix \mathbf{R} is given by

$$\mathbf{R} = R_x^{s_k} \otimes R_x^{s_l} + R_y^{s_k} \otimes R_y^{s_l}, \quad (18)$$

where \otimes denotes the Kronecker product, and with

$$R_x^{s_k}(k, k') = -\frac{1}{s_k} \frac{\partial^2 B^{2n+1}}{\partial x^2}, (k' - k) \quad R_y^{s_k}(k, k') = s_k B^{2n+1}(k' - k),$$

$$R_x^{s_l}(l, l') = s_l B^{2n+1}(l' - l), \quad R_y^{s_l}(l, l') = -\frac{1}{s_l} \frac{\partial^2 B^{2n+1}}{\partial y^2}(l' - l), \quad (19)$$

with $k, k' = 1, 2, \dots, N_k$ and $l, l' = 1, 2, \dots, N_l$.

The Minimizer for the $SE(2)$ Case in Matrix-Vector Notation. For the $SE(2)$ case, the shape of the energy functional $\mathcal{E}(T)$ and the corresponding minimizer are the same as for $E(t)$ on the \mathbb{R}^2 case, and are given by (14) and (15). However, the definitions of S , R and \mathbf{c} are different. In this case S is a $[P \times N_k N_l N_m]$ matrix given by

$$S = \left\{ (s_{1,1,1}^i, s_{1,1,2}^i, \dots, s_{1,1,N_m}^i, s_{1,2,1}^i, \dots, s_{1,2,N_m}^i, \dots, s_{1,N_l,N_m}^i, \dots, s_{N_k,N_l,N_m}^i) \right\}_{i=1}^P, \\ s_{k,l,m} = (B_{s_k s_l s_m}^n * (M \hat{U}_{f_i}))(k, l, m), \quad (20)$$

with $B_{s_k s_l s_m}^n(x, y, \theta) = B^n\left(\frac{x}{s_k}\right) B^n\left(\frac{y}{s_l}\right) B^n\left(\frac{\theta \bmod 2\pi}{s_m}\right)$, with angular resolution parameter $s_m = 2\pi/N_m$. Vector \mathbf{c} is a $[N_k N_l N_m \times 1]$ column vector containing the B-spline coefficients and is stored as follows:

$$\mathbf{c} = (c_{1,1,1}, c_{1,1,2}, \dots, c_{1,1,N_m}, c_{1,2,1}, \dots, c_{1,2,N_m}, \dots, c_{1,N_l,N_m}, \dots, c_{N_k,N_l,N_m})^T. \quad (21)$$

The explicit expression of $[N_k N_l N_m \times N_k N_l N_m]$ matrix R is given in Appendix A.

3 Optic Nerve Head Detection

3.1 Processing Pipeline

The location of the ONH is found through the following five steps:

1. The input image f is (locally) normalized via the Luminosity-Contrast normalization method described by Foracchia et al. [18], giving f_{lcn} .
2. To further reduce sensitivity to high intensity structures, we apply the following intensity mapping $f_{lcn}^{erf} = \text{erf}(8 f_{lcn})$, with $\text{erf}(i) = \frac{2}{\sqrt{\pi}} \int_0^i e^{-x^2} dx$ the error function. The effect is a soft binarization of the image, by which more emphasis is put on contextual information rather than intensity information.
3. In case of a 2D template t , \hat{f}_{lcn}^{erf} is approximated by (5). In case of an orientation score template T , the score $U_{f_{lcn}^{erf}}$ is calculated via (1), and is normalized after taking the modulus giving $|\hat{U}|_{f_{lcn}^{erf}}$.
4. An ONH probability map $P^t(g) = \hat{c}_{t,f}(g)$ is calculated via (4a), or $P^T(g) = \hat{C}_{T,U_f}(g)$ is calculated via (6) in case of an orientation score template.
5. An ONH probability map constructed using template τ (with $\tau = t$ or $\tau = T$) is denoted by P^τ . In case multiple templates are used, each probability map P^τ is rescaled to a range of $[0, 1]$. The final optic nerve head location is then calculated as $g_o = \underset{g \in SE(2)}{\text{argmax}} \left(\sum_{\tau \in \mathcal{T}} P^\tau(g) \right)$, with \mathcal{T} the set of templates used.

Since the ONH generally appears under the same orientation in every image, we restrict our search for the ONH location $g_o = (\mathbf{x}_o, \theta_o)$ to translations \mathbf{x}_o only, and assume $\theta_o = 0$. To reduce computation time the image is rescaled by a factor of $\frac{r_{target}}{r_{est}}$, with $r_{target} = 20$ pixels and r_{est} the estimated optic disk radius. For normalization in step 1 we used a window size of $\frac{1}{2} r_{target}$, for the orientation score transforms we used cake wavelets [4] with angular resolution $s_\theta = \frac{\pi}{12}$. For normalization we have used isotropic mass functions $M(\mathbf{x}, \theta) = m(\mathbf{x})$, for details see [14, Section 2.4].

3.2 Templates

In our experiments we have considered three different types of templates: model templates, average templates and trained templates. In total we will be investigating 6 different templates, labeled **A-F**, see Table. 1. Template **A** is a disk filter and models the shape of the optic disk, and has been used for ONH detection in [8]. Template **B** is a template that models the pattern of blood vessels radiating outwards from the ONH. This template is described in our previous work [14]. Templates **C** and **D** are average templates, and are respectively found by $t_C = \frac{1}{P} \sum_{i=1}^P f_i$ and $T_D = \frac{1}{P} \sum_{i=1}^P |U_{f_i}|$, with $\{f_1, \dots, f_P\}$ the set of positive ONH image patches, and $\{U_{f_1}, \dots, U_{f_P}\}$ the set of orientation scores hereof.

Templates **E-F** are trained using the methods described in Subsection 2.3. The number of B-splines was set to $N_k = N_l = 50$ and $N_m = 12$. Template **E** is constructed in the \mathbb{R}^2 domain with regularization parameter $\lambda = 10^{-1.5}$. The orientation score template **F** is constructed using regularization parameters $\lambda = 10$ and $D_{\theta\theta} = 10^{-3.5}$.

Templates **C-F** require a training set. The set is constructed using the first $P = 100$ images of the publicly available MESSIDOR database (<http://messidor.crihan.fr/index-en.php>). Each positive optic nerve head patch f_i (with label $y_i = 1$) is centered at the ONH and has a square window size of $8 r_{target}$. The negative patches were selected based on the critical areas for template **C**. Each negative patch f_i (with label $y_i = 0$) is centered around the largest local maximum of the image filtered with template **C**, and which does not lie within the circumference of the optic disk. See Fig. 2 for examples of positive and negative patches. The images used in the training underwent the same first three processing steps as described in Subsection 3.1. For processing of conventional (RGB) fundus images we used the green channel. For SLO images we used the near-infrared color channel of the first $P = 100$ images of our private SLO image database, which will be described in the next section.

4 Results and Discussion

Data. For validation, we made use of a private database consisting of 208 SLO images taken with an EasyScan (i-Optics B.V., the Netherlands) and 208 CF images taken with a Topcon NW200 (Topcon Corp., Japan). For full details see [14]. The two sets of images are labeled as "ES" and "TC" respectively. Our method is also tested on three widely used public databases: MESSIDOR, DRIVE (<http://www.isi.uu.nl/Research/Databases/DRIVE>) and STARE (<http://www.ces.clemson.edu/~ahoover/stare>), consisting of 1200, 40 and 81 images respectively. For each image, the detected ONH position was marked as correct if it was located within the circumference of the actual ONH. To this end we used the annotations kindly provided by the authors of [6] (<http://www.uhu.es/retinopathy>), and manually outlined the ONH border for the other databases.

Results and Discussion. Results of our ONH detection framework are given in Table 1. The results for single template methods are categorized in three

Table 1. Results of (combinations) of templates for optic nerve head detection (number of fails in parentheses).

Template ID	Domain	ES (SLO) 208	TC 208	MESSIDOR 1200	DRIVE 40	STARE 81	All Images 1737
model templates							
A	\mathbb{R}^2	65.38% (72)	95.19% (10)	82.00% (216)	87.50% (5)	53.09% (38)	80.37% (341)
B	$SE(2)$	62.50% (78)	77.40% (47)	86.67% (160)	67.50% (13)	65.43% (28)	81.23% (326)
average templates							
C	\mathbb{R}^2	99.52% (1)	99.04% (2)	98.00% (24)	95.00% (2)	67.90% (26)	96.83% (55)
D	$SE(2)$	99.52% (1)	100.0% (0)	99.50% (6)	97.50% (1)	93.83% (5)	99.25% (13)
trained templates							
E	\mathbb{R}^2	92.79% (15)	98.56% (3)	94.58% (65)	92.50% (3)	50.62% (40)	92.75% (126)
F	$SE(2)$	100.0% (0)	98.56% (3)	99.67% (4)	100.00% (0)	90.12% (8)	99.14% (15)
combinations of two templates							
D + F		100.0% (0)	100.0% (0)	99.75% (3)	100.0% (0)	97.53% (2)	99.71% (5)
C + D		100.0% (0)	100.0% (0)	99.58% (5)	100.0% (0)	85.19% (12)	99.02% (17)
C + F		100.0% (0)	100.0% (0)	99.58% (5)	100.0% (0)	85.19% (12)	99.02% (17)
E + F		100.0% (0)	100.0% (0)	99.50% (6)	100.0% (0)	83.95% (13)	98.91% (19)
E + D		100.0% (0)	100.0% (0)	99.42% (7)	100.0% (0)	81.48% (15)	98.73% (22)
...							

categories: model templates, average templates and trained templates. For combinations of two templates only the best five combinations are shown.

Firstly, we observe that templates acting in the domain $SE(2)$ of an orientation score considerably outperform their 2D equivalents. The orientation score templates put more emphasis on the pattern of blood vessels, rather than intensity features, and are therefore more robust against bright lesions and other pathologies. The advantage of our extension to $SE(2)$ is best observed on the challenging STARE database, which contains a wide variety of severely pathological images.

Secondly, from Table 1 we see that average as well as trained templates outperform basic model templates, with the average templates slightly outperforming the trained templates. While the two best templates **D** and **F** individually give excellent performances, an even higher performance can be achieved in combining templates. With an accuracy of 99.71%, only 5 fails out of 1737 images, this combination **D+F** outperforms all other combinations of the templates used in this paper. Here we stress the crucial role of our proposed template optimization scheme that provides additional means for the construction of complementary templates; with single templates alone such high accuracy could not have been achieved. Exemplary results of matching with templates **D** and **F** are shown in in Fig. 4.

Thirdly, we note that when combining templates it is favorable to stay within the $SE(2)$ framework. From the results we see that detection with \mathbb{R}^2 -type templates is improved by combination with $SE(2)$ -type templates. However, with respect to the use of single $SE(2)$ -type templates, these combinations are not favorable (compare e.g. the result of **C**, **D** and **C+D**). In future work we will therefore put more focus on the design of complementary $SE(2)$ -type templates. As such, we will investigate the ability of regularization parameter $D_{\theta\theta}$ to tune



Fig. 4. A selection of successful ONH detections in challenging images.

the template to anisotropic or isotropic structures (see Fig. 3) as a means to construct different/complementary $SE(2)$ templates.

Finally, in Table 2 we compare our method to the state of the art on ONH detection (for full comparison to literature see [11, 14, and references therein]). Although our correlation-based method is rather basic in nature, it competes well with the state of the art. Only the method by Lu slightly outperforms our method by ones less false detection on the STARE database. Furthermore, as our detection framework merely relies on a sequence of correlations the method is highly parallelizable for speed optimization. Our current implementation (in Python) of the entire processing pipeline, including preprocessing (see Subsection 3.1), takes on average 1.1s per image.

Table 2. Comparison to state of the art: Optic nerve head detection results (number of fails in parentheses). The most recent five methods were selected for comparison. For a full comparison to literature see [11].

Database	Size	Proposed	Ramakanth et al. [11]	Yu et al. [8]	Lu et al. [7]	Lu. [9]	Mahfouz et al. [10]
DRIVE	40	100.0% (0)	100.0% (0)	-	97.5% (1)	-	100.0% (0)
STARE	81	97.53% (2)	93.83% (5)	-	96.3% (3)	98.77% (1)	92.59% (6)
MESSIDOR	1200	99.75% (3)	99.42% (7)	99.0% (12)	-	99.75% (3)	-
Av. time (s)		1.1	0.21	4.7	40	5	0.65

5 Conclusion

In this paper we have extended the concept of object detection via (normalized) cross correlation in the image domain \mathbb{R}^2 , to the domain $\mathbb{R}^2 \times S^1$ of orientation scores. The extension allows for the efficient detection of orientation patterns, while staying in the intuitive and efficient framework of template matching via cross correlation. Furthermore we have presented a method for the construction of templates to be used in this matching framework. The method was tested in the application to optic nerve head detection in retinal images. Here we achieved a success rate of 99.71% on a set of 1737 images, with an average processing time of 1.1s per image. The method is generically applicable, and is especially beneficial for the detection of objects characterized by orientated/line structures.

Acknowledgements: This work is part of the Hé Programme of Innovation, which is (partly) financed by the Netherlands Organisation for Scientific Research (NWO). Also, the research leading to these results has received funding from the ERC council under the EC's 7th Framework Programme (FP7/2007–2013) / ERC grant agr. No. 335555.

References

1. Lewis, J.: Fast normalized cross-correlation. *Vision Interface* **10**(1) (1995) 120–123
2. Yoo, J.C., Han, T.: Fast normalized cross-correlation. *CSSP* **28**(6) (2009) 819–843
3. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Volume 1. (June 2005)* 886–893 vol. 1
4. Duits, R., Felsberg, M., Granlund, G.H., ter Haar Romeny, B.M.: Image analysis and reconstruction using a wavelet transform constructed from a reducible representation of the euclidean motion group. *IJCV* **72**(1) (2007) 79–102
5. Sinthanayothin, C., Boyce, J.F., Cook, H.L., Williamson, T.H.: Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images. *The British Journal of Ophthalmology* **83**(8) (1999) 902–10
6. Aquino, A., Gegundez-Arias, M.E., Marin, D.: Detecting the optic disc boundary in digital fundus images using morphological, edge detection, and feature extraction techniques. *IEEE TMI* **29**(11) (2010) 1860–9
7. Lu, S., Lim, J.: Automatic optic disc detection from retinal images by a line operator. *Biomedical Engineering, IEEE TBME* **58**(1) (2011) 88–94
8. Yu, H., et al.: Fast localization and segmentation of optic disk in retinal images using directional matched filtering and level sets. *IEEE TITB* **16**(4) (2012) 644–57
9. Lu, S.: Accurate and efficient optic disc detection and segmentation by a circular transformation. *IEEE TMI* **30**(12) (Dec 2011) 2126–2133
10. Mahfouz, A., Fahmy, A.: Fast localization of the optic disc using projection of image features. *IEEE TIP* **19**(12) (Dec 2010) 3285–3289
11. Ramakanth, S.A., Babu, R.V.: Approximate nearest neighbour field based optic disk detection. *Computerized Medical Imaging and Graphics* **38**(1) (2014) 49 – 56
12. Jonas, J.B., Budde, W.M., Panda-Jonas, S.: Ophthalmoscopic Evaluation of the Optic Nerve Head. *Survey of Ophthalmology* **43**(4) (1999)
13. Hubbard, L.D., et al.: Methods for evaluation of retinal microvascular abnormalities associated with hypertension/sclerosis in the Atherosclerosis Risk in Communities Study. *Ophthalmology* **106**(12) (December 1999) 2269–80
14. Bekkers, E., Duits, R., ter Haar Romeny, B.: Optic nerve head detection via group correlations in multi-orientation transforms. In Campilho, A., Kamel, M., eds.: *Image Analysis and Recognition. LNCS. Springer* (2014) 293–302
15. Freeman, W., Adelson, E.: The design and use of steerable filters. *IEEE TPAMI* **13**(9) (Sep 1991) 891–906
16. Citti, G., Sarti, A.: A cortical based model of perceptual completion in the roto-translation space. *JMIV* **24**(3) (2006) 307–326
17. Duits, R., Franken, E.: Left-invariant parabolic evolutions on $SE(2)$ and contour enhancement via invertible orientation scores part I: Linear left-invariant diffusion equations on $SE(2)$. *Quarterly of Applied Mathematics* **68**(2) (2010) 255–292
18. Foracchia, M., Grisan, E., Ruggeri, A.: Luminosity and contrast normalization in retinal images. *MEDIA* **9**(3) (2005) 179–90

A Explicit Expression of the Regularization Matrix

Matrix R for regularization of templates in $SE(2)$, see Section 2.3, is given by

$$R = D_{11}R_\xi + D_{22}R_\eta + D_{33}R_\theta, \quad (22)$$

with regularization matrix

$$\begin{aligned} R_\xi = & \left(R_\xi^{I s_k} \otimes R_\xi^{I s_l} \otimes R_\xi^{I s_m} \right) + \left(R_\xi^{II s_k} \otimes R_\xi^{II s_l} \otimes R_\xi^{II s_m} \right) \\ & + \left(R_\xi^{III s_k} \otimes R_\xi^{III s_l} \otimes R_\xi^{III s_m} \right) + \left(R_\xi^{IV s_k} \otimes R_\xi^{IV s_l} \otimes R_\xi^{IV s_m} \right) \end{aligned} \quad (23)$$

of which the elements of the matrices are given by

$$\begin{aligned} R_\xi^{I s_k}(k, k') &= -\frac{1}{s_k} \frac{\partial^2 B^{2n+1}}{\partial u^2}(k' - k), & R_\xi^{I s_l}(l, l') &= s_l B^{2n+1}(l' - l), \\ R_\xi^{I s_m}(m, m') &= \int_0^\pi \cos^2(\theta) B^n\left(\frac{\theta}{s_m} - m\right) B^n\left(\frac{\theta}{s_m} - m'\right) d\theta, \\ R_\xi^{II s_k}(k, k') &= -R_\xi^{III s_k}(k, k') = \frac{\partial B^{2n+1}}{\partial v}(k' - k), \\ R_\xi^{II s_l}(l, l') &= -R_\xi^{III s_l}(l, l') = -\frac{\partial B^{2n+1}}{\partial v}(l' - l), \\ R_\xi^{II s_m}(m, m') &= R_\xi^{III s_m}(m, m') = \int_0^\pi \cos(\theta) \sin(\theta) B^n\left(\frac{\theta}{s_m} - m\right) B^n\left(\frac{\theta}{s_m} - m'\right) d\theta, \\ R_\xi^{IV s_k}(k, k') &= s_k B^{2n+1}(k' - k), & R_\xi^{IV s_l}(l, l') &= -\frac{1}{s_l} \frac{\partial^2 B^{2n+1}}{\partial v^2}(l' - l), \\ R_\xi^{IV s_m}(m, m') &= \int_0^\pi \sin^2(\theta) B^n\left(\frac{\theta}{s_m} - m\right) B^n\left(\frac{\theta}{s_m} - m'\right) d\theta, \end{aligned} \quad (24)$$

with regularization matrix

$$\begin{aligned} R_\eta = & \left(R_\xi^{II s_k} \otimes R_\xi^{II s_l} \otimes R_\xi^{IV s_m} \right) - \left(R_\xi^{II s_k} \otimes R_\xi^{II s_l} \otimes R_\xi^{II s_m} \right) \\ & - \left(R_\xi^{III s_k} \otimes R_\xi^{III s_l} \otimes R_\xi^{III s_m} \right) + \left(R_\xi^{IV s_k} \otimes R_\xi^{IV s_l} \otimes R_\xi^{I s_m} \right), \end{aligned} \quad (25)$$

and with regularization matrix

$$R_\theta = (R_\theta^{s_k} \otimes R_\theta^{s_l} \otimes R_\theta^{s_m}), \quad (26)$$

of which the elements of the matrices are given by

$$\begin{aligned} R_\theta^{s_k}(k, k') &= s_k B^{2n+1}(k' - k), \\ R_\theta^{s_l}(l, l') &= s_l B^{2n+1}(l' - l), & R_\theta^{s_m}(m, m') &= -\frac{1}{s_m} \frac{\partial^2 B^{2n+1}}{\partial w^2}(m' - m). \end{aligned} \quad (27)$$

Note that the four separate terms $I-IV$ of Eq. (23) arise from the left invariant derivative ∂_ξ : $\left| \frac{\partial T}{\partial \xi} \right|^2 = \left| \cos(\theta) \frac{\partial T}{\partial x} + \sin(\theta) \frac{\partial T}{\partial y} \right|^2$.