Automatic detection of the optic disc, fovea and vascular arch in digital color photographs of the retina

Meindert Niemeijer, Bram van Ginneken
Image Sciences Institute
University Medical Center Utrecht
Utrecht, The Netherlands
meindert@isi.uu.nl

Frank ter Haar
IICS, Utrecht University
Utrecht, the Netherlands

Michael D. Abràmoff
Dept. of Ophthalmology and Visual Sciences
University of Iowa, Iowa, USA

Abstract

We present a novel method that determines whether a macula centered retinal image is from the left or right eye and automatically detects the optic disc, the fovea and the vascular arch by inferring the location of a set of landmarks placed on these structures. The algorithm relies on a specific energy function that combines global and local cues. The global cues are derived from vascular atlases of the vessel orientation and thickness on the retina as well as a vascular distance map. A fourth component models the local appearance around each of the landmarks in the model and is able to estimate the distance between a position in the image and the target position of a landmark. For the minimization of the energy function a combination of optimization methods is used. We compare the results of several different system setups and combinations of energy function components with the performance of a second human observer. The best performing system localizes the OD in 91% of all cases, the fovea in 94% of all cases and correctly positions 74% of all vessel landmarks. The results show that a combination of global and local energy function components is required to obtain optimal results.

1 Introduction

The automatic determination of the position of distinctive points is an important and recurrent theme in medical image analysis. Knowledge about the position of such landmarks can be used for many purposes, e.g. to initialize segmentations or to register a group of images. In this work we are concerned with the identification of three important structures in retinal images, namely the optic disc, the fovea and the vascular arch (see Figure 1), which we denote with a set of 16 points. We pursue a supervised approach to the automatic determination of the location of these points in new (unseen) images, that is, we use a set of images in which these points are provided beforehand to train our algorithm.

BMVC 2005 doi:10.5244/C.19.17

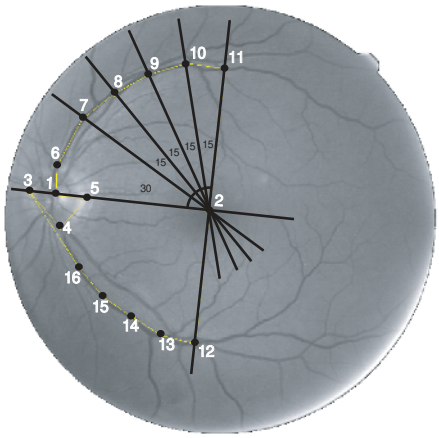

Figure 1: The model overlaid on a retinal image. The first two landmarks, 1 located in the middle of the OD and 2 located on the fovea, define the main axis of the model. Landmarks 3-6 are located on the edge of the optic disc. The remaining landmarks, 7 to 16, define the venous vascular arch.

There exist several supervised algorithms for locating landmarks in images. Well-known examples are active shape models [4], active appearance models [2] and active blobs [14]. A key element of these approaches is the modelling of gray-level information: for a given image and a set of point positions, a function evaluates how well the points at their given locations 'fit' to the image. Active blobs and active appearance models use a single *global* measure to determine the fit, derived from all pixel values in the convex hull of the model. Active shape models, on the other hand, employ a collection of *local* models around each landmark. We argue that in many tasks, including the retinal image analysis application considered here, a single fit function of a predetermined type is not sufficient. Instead, a dedicated fit function is required that consists of both global and local terms, and that should be specifically designed for the task at hand. The key contribution of this work is to describe such a function for retinal images, to outline a general framework in which this function can be used to determine the landmark positions and to evaluate the algorithm on a large set of retinal image data.

2 Previous Work

The closest related previous work was done by Li et al. [9], who presented a model based approach which starts by detecting the OD location. Then an active shape model is used to extract the main course of the vasculature. This is followed by the fitting of a parabolic model to the vasculature, and detection of the fovea. A full quantitative evaluation is not given. All other previous work focuses on just one of the three tasks considered here.

3 Method

3.1 General Outline of the Method

The output of the algorithm consists of a vector **s** which contains the positions of a set of n distinct points (landmarks) $(x_i, y_i), i = 1 \ldots n$ in a previously unseen image I. With this output, a set of images can be co-registered by warping them to a common frame of reference, typically given by the mean shape $\bar{\mathbf{s}}$. Using this warping function (we use a warping algorithm based on radial basis functions [13]) any position in I can be translated to a position in the common frame of reference.

For training, a set of images must be provided for which **s** is given. The algorithm requires two models. One is a shape model, which generates a shape **s** from a set of shape parameters **b**. There are thus two spaces, the shape space in which **b** lives and the image space in which **s** resides. Typically the dimensionality and extent of the shape space is far smaller than that of the image space; every **b** should yield a valid shape **s**, but not every configuration of n points located somewhere in the image (space) is a valid shape. The second model is an energy function $F(\mathbf{s}, I)$ that returns how well **s** fits to I. The algorithm must find the shape parameters **b** that minimize F. In practice, F will have many local minima and finding the global minimum is a difficult problem.

There are two approaches to minimizing F. The standard approach would be to choose a standard optimization algorithm that searches the shape space. Alternatively, one could move the points (x_i, y_i) individually in the image space and fit the shape model to the resulting **s**. This requires that the shape model can be inverted, i.e. **b** can be computed given **s**.

A proper choice for F is essential. A particular application will often suggest multiple ways to measure the fitness of a solution and thus F may consist of several terms. These terms can be global, in the sense that the complete image or a large part of it are used to evaluate the term, or local, in that they require only one landmark and its neighborhood in the image. It is also possible to have terms that do not depend on the image but only on **b** – typically a penalty term that favors only likely shapes.

In our application, F consists of a combination of local and global terms. The minimization algorithm operates both in the shape and image space. The next subsections describe the shape model, energy function and optimization strategy employed in our application.

3.2 Shape Model

We defined a set of 16 points covering the major anatomical retinal landmarks (see Figure 1). Several constraints were introduced in the localization of the landmarks. Landmarks 1 and 2 define the main axis of the model. All the landmarks on the vascular arch (7-16) are positioned at a fixed angle from this main axis. Two of the four landmarks (3 and 5) on the border of the OD are defined to lie on the main axis while the other two (4 and 6) should be located at the point where the venous vascular arch leaves the OD.

The shape model itself was obtained by computing the mean shape and the covariance

$$\bar{\mathbf{s}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{s}_i \qquad \mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{s}_i - \bar{\mathbf{s}})(\mathbf{s}_i - \bar{\mathbf{s}})^T \tag{1}$$

from a set of 100 training cases, and retaining the t eigenvectors $\mathbf{v}_i, i = 1 \ldots t$ with the largest eigenvalues $\lambda_1 \ldots \lambda_t$ from $\mathbf{S}$ which span the shape space. Shape parameters b_i are defined as the projections on $\mathbf{v}_i$ scaled by $\sqrt{\lambda_i}$ so that b_i should typically lie in a range $[-3, 3]$ standard deviations. This simple linear type of shape model is known as a Point Distribution Model (PDM) [4]. The shapes $\mathbf{s}$ are not aligned prior to the construction of the shape model and thus the variation in location, rotation and scaling of the points is contained in $\mathbf{b}$. We choose $t = 8$, in which case the PDM explains 96.45% of the variation in the training set.

3.3 Energy Function

The global terms in the proposed energy function are all derived from a map (an atlas) of the vasculature of the retina. The vasculature in macula centered retinal images consists of two major branches, an upper and a lower branch, which together form the vascular arch (see Figure 1). Both branches consist of at least one major artery and at least one major vein. Usually the vein is the thickest and darkest vessel in the branch. This described global vascular pattern is always present on the retina.

The vascular arch defines the position of the OD which is located at its apex where the upper and lower branch meet. The fovea is approximately at a fixed offset from the OD and in an area devoid of vessels. Thus the locally dominant orientation of vessels, the width of the vessels and vessel density (the distance to nearby vessels) all provide important cues to locate these landmarks.

3.3.1 Vessel Thickness and Orientation Atlas

The general use of image orientation for model fitting has been proposed before by Cootes [3]. The idea to use the global orientation of the retinal vasculature to locate a point on the retina was suggested by Foraccia et al. [7]. They proposed a mathematical model of the retinal vessel orientation which, after fitting, produced an estimate of the OD location. Because of the variability in the vessel orientation In contrast we segment the vasculature to construct an atlas which maps the average orientation as well as the average vessel thickness across the image.

These atlases are constructed by a four step process: First, the vasculature in all training images is segmented and the vessel centerline pixels are found. Second, the thickness t_i and orientation θ_i for each centerline pixel i is determined. Third, the centerlines are warped into a common frame of reference, given by the mean shape. Fourth, for each pixel an average thickness and orientation is determined.

The vasculature was segmented using a pixel classification based method we previously developed [11]. This technique assigns a posterior probability to each pixel in an image which indicates whether the pixel is part of a vessel. A threshold can be applied to obtain a binary vessel segmentation.

Next, the binary vessel segmentation is thinned [6] to obtain the approximate vessel centerlines. These are disconnected at the branching points (pixel has more than two neighbors) and small segments (≤ 5 pixels) are removed.

For each centerline pixel i the orientation θ_i is determined by applying PCA on the seven connected centerline pixels centered on i. The vessel thickness t_i at i is defined as the distance between the edges of the vessel at i.

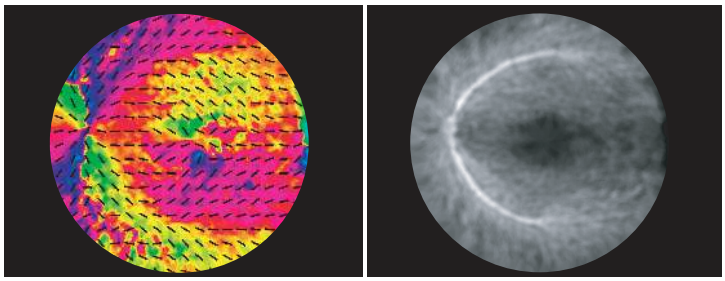

Figure 2: The image on the left shows the orientation atlas, the local orientation is visualized with black bars. All orientations have been mapped to the Hue circle of the HSI-color space. On the right the vessel thickness atlas is shown.

To form the atlas, the centerline pixels of the images of the training set are warped into a common frame of reference given by the mean shape $\bar{\mathbf{s}}$.

The final step is to calculate the average thickness and orientation for every centerline pixel i in the atlas (see Figure 2). Any remaining "holes" in the atlas are filled with a nearest neighbor averaging scheme. Every pixel in the atlas is assigned the mean value of its $k = 121$ nearest neighbors (see Figure 2b). This value for k was determined after preliminary experiments but is not a critical parameter.

To determine the fitness of a new shape $\mathbf{s}$ generated by the PDM on a new image I, step 1 and 2 of the discussed scheme are first applied to I. Then, whenever the fit measures for a new $\mathbf{s}$ are needed, step 3 with $\mathbf{s}$ is used. The fitness value $f_{thickness}$ and $f_{orientation}$ are calculated by

$$f_{thickness} = \sum_{i=1}^{n} |t_i - t_i^{atlas}| \qquad f_{orientation} = \sum_{i=1}^{n} t_i |\theta_i - \theta_i^{atlas}| \tag{2}$$

where n is the total number of centerline pixels, t_i^{atlas} is the thickness value of pixel i in the atlas and θ_i^{atlas} is the orientation of pixel i in the atlas.

3.3.2 Sum of Inverted Distance Maps (SIDM)

The vessel landmarks (7-16) are often located on or near the thickest vessels in the vascular arch. The SIDM procedure will yield an image I_{SIDM} in which the pixels on and close to the thickest vessels will have higher values than pixels near to thinner vessels.

Given a binary vessel segmentation (see Section 3.3.1) I_{IDM} is produced in two steps, first a distance transform is applied, followed by an inversion (multiplication with -1). Next, the binary vessel segmentation is eroded by eliminating all boundary pixels. These two steps are repeated iteratively until all vessel pixels have been removed. After each iteration I_{IDM} is stored. Now the result image I_{SIDM} and the fitness value f_{SIDM} are calculated by:

$$I_{SIDM} = \frac{1}{N} \sum_{i=1}^{N} I_{IDM} \qquad f_{SIDM} = -\sum_{i=1}^{p} v_i \tag{3}$$

where N is the number of I_{IDM} and where p iterates over the 10 vessel landmarks in the PDM and v_i is the pixel value in I_{SIDM} at the position of landmark i.

3.3.3 Pixel Distance Regression

In addition to the previous global terms, a term which depends on the local gray value appearance around the landmarks in the training set was also used. The pixel distance regression models the local gray-value information at different distances d in a circular region of interest ROI_p around each of the landmarks p. Then given a pixel from a new image, the distance from the pixel to any p can be estimated.

The size of ROI_p is determined by its radius r. The value of r is dependent on the scale of the object indicated by p. For example, when p is located on a vessel, r should be smaller than when p is positioned on the fovea. We distinguish three different ROI_p; vessel, OD and fovea. The following settings for r were determined in preliminary experiments; $r = 10$ for the vessel landmarks, $r = 15$ for the OD landmarks and $r = 25$ for the fovea landmark.

To learn the local appearance around each of the manually indicated p in the training set, pixels in ROI_p are represented by a rich set of local features, namely the output of a filterbank consisting of the Gaussian filter and its 1st and 2nd derivatives (i.e. $L, L_x, L_y, L_{xx}, L_{xy}, L_{yy}$) at multiple scales $\sigma = 1, 2, 4, 8, 16, 20$.

After sampling every fourth pixel from ROI_p in every training image, all samples are normalized to zero mean and unit standard deviation. Now, given a new pixel i we use k-nearest neighbor (kNN) regression to estimate the distance to the "correct" position of landmark p by taking the median value of d for the k nearest neighbors in feature space. The implementation of the kNN algorithm was developed by Arya et al. [1], it allows for fast approximate neighbor retrieval. We set parameter $k = 5$.

The final fit value f_{dist} is calculated by

$$f_{dist} = \sum_{p=1}^{n} d_p \tag{4}$$

where n is the number of landmark points in the model and d_p is the estimated distance from the correct position for landmark point p.

3.3.4 Component Combination

Before doing a linear combination of the four components their outputs are first normalized to have a value between 0 and 1. The final output value then is

$$F(\mathbf{s}, I) = \omega_1 f_{thickness} + \omega_2 f_{orientation} + \omega_3 f_{SIDM} + \omega_4 f_{dist} \tag{5}$$

where ω_i represents the weight associated with a component, set to either 0 or 1 in our experiments but can be set to any value.

3.4 Optimization

We evaluated three standard optimization techniques that operate in the shape space; Simulated Annealing (SA) [8, 5], Simplex optimization [10] and Powell's optimization method [12]. The settings for SA were chosen such that its runtime did not exceed the maximally allowed reasonable runtime of 10 minutes. As the Simplex algorithm was much faster (average runtime 35 seconds) it was started six times with different starting positions. Powell was always ran once, starting from the mean shape in the shape space.

Table 1: Overview of the different configurations that were run on the test set.

System nr.	1	2	3	4	5	6
Global cues	×	×	×	×	-	-
Local cues	×	×	-	-	×	×
Image space optimization	×	-	×	-	×	-

The availability of the vessel thickness atlas, and the observation that the landmarks 7 to 16 are constrained to move along straight lines through landmark 2 suggested a simple procedure to move these points individually. Lines are projected from landmark 2 through each of the vessel landmarks. Along these lines the thickest vessel with the highest contrast is found and the landmark is positioned there. The local contrast of a vessel is easily determined using the available vessel segmentation and local orientation θ_i of the vessel centerline.

The described image space operation was integrated into the optimization strategy as follows. First, one run of the optimization algorithms was executed, then vessel landmarks were relocated using the previously described procedure. Then, shape $\mathbf{s}$, the result after fitting the PDM, was used as the starting position in shape space for a new run of the optimization algorithm. If the value of F was lower for the shape produced after the second run it was considered the final result, otherwise the result before image space optimization was kept.

4 Experiments and Results

In the development of the system a total of 220 different macula centered digital color fundus photographs were used. They were randomly selected from a set of photos obtained from a diabetic retinopathy screening program in The Netherlands. Each image is captured using 8 bits per color plane has a circular field of view of 540 pixels diameter. The green plane of each image is used.

These 220 images were divided among a training set of 100 images, a test set of 100 images and a separate development set of 20 images reserved for preliminary experiments. Although both a patient's left and right eye could be included in one set, data of a single patient was never spread among the different sets. In the training and development set all images were mirrored when necessary to resemble photographs of the left eye (OD on the left side of the image). The test set was left unprocessed.

In each of the images a reference standard was established by letting an observer, the first author, mark sixteen points (see Figure 1). Then, the test set was annotated again by a second observer, the second author.

In preliminary experiments, different optimization algorithms and several parameter settings were tested. The best results were obtained with simplex optimization in combination with the image space optimization procedure. A total of six different systems (see Table 1), all using simplex optimization, were run on the test set.

To distinguish between left and right eyes in the test set the model was fitted two times to every image (normal and mirrored). The best model fit indicated the type of eye.

Table 2: Performance overview of the different systems on the test set. From left to right, successful OD segmentation, OD center error (1), found foveae landmarks, fovea landmark error (2), found vessel landmarks, vessel landmarks error (7-16) and the error for the whole shape. All errors are averages and are always given in pixels.

System	OD detected	OD error	Fovea detected	Fovea error	Vessel detected	Vessel error	Total error
1	91.0%	10.1	94.0%	10.2	74.0%	12.9	11.8
2	88.0%	11.2	94.0%	9.8	63.0%	17.4	14.8
3	62.0%	20.0	71.0%	16.8	43.3%	23.2	21.5
4	80.0%	14.3	77.0%	14.7	56.4%	19.1	17.1
5	84.0%	11.0	96.0%	9.2	69.0%	16.2	14.2
6	81.0%	12.4	97.0%	9.6	40.5%	24.3	19.8
2^{nd} observer	97.0%	4.9	99.0%	4.6	97.0%	5.9	5.36
Mean shape	59.0%	18.5	97.0%	11.2	26.6%	28.3	24.1

4.1 Results

Table 2 summarizes the results. We determined that all the presented systems were able to distinguish left from right eyes for all images in the test set.

Performance was measured separately for each of the three groups of landmarks (i.e. OD, vessel and fovea). For the OD points, both overlap as well as the difference in the localization of the OD center were chosen as the evaluation measures. We define that an OD localization is successful if there is at least 55% overlap. The OD surface was estimated by fitting a circle through landmarks 3-6. Vessel and fovea points are considered successfully placed if they are located within a certain distance from the reference standard. This distance is determined for each landmark separately, and is defined as the maximum distance between the reference standard and the second observer. Before measuring the maximum distance all clear outliers (e.g. cases in which the second observer selected a different vessel) were excluded.

5 Conclusion & Discussion

A new method has been presented that determines whether a macula centered retinal image is a left or right eye and automatically detects the optic disc, the fovea and the vascular arch by inferring the location of a set of landmarks placed on these structures. The algorithm relies on a specific energy function that combines global and local cues. The results show that a combination of global and local energy function components is needed to obtain the best system performance. Also the success of a combined optimization method in both the shape space as well as the image space is evident from the results.

Figure 3 shows the model fitted to three images. One successful case and two cases demonstrating typical errors. The middle image illustrates a case in which the average pixel distance between the fit and the reference standard will be large, but the generated shape seems a correct “alternative” to the reference standard. In the rightmost image a more serious error is shown, where the vascular arch was found but due to an incorrect

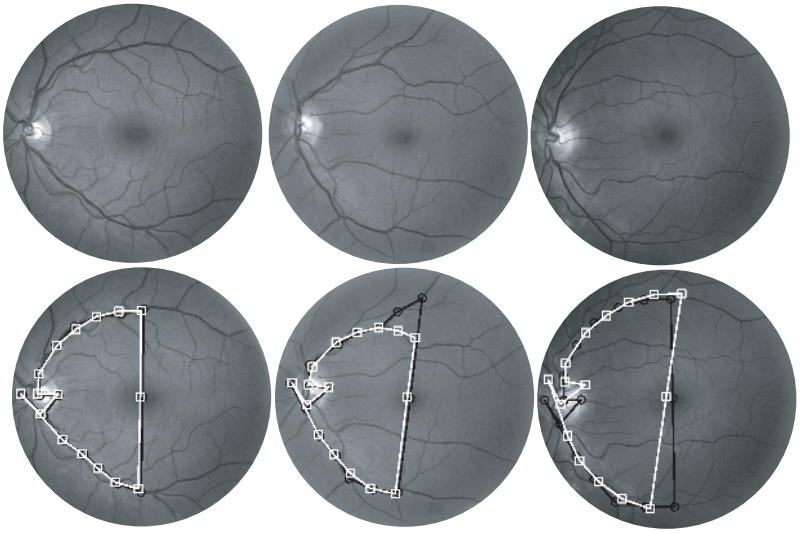

Figure 3: Three example model fits, with the original images at the top and the images with both the model fit (white) and the reference standard (black) on the second row. The left image illustrates a good result, the other two images demonstrate typical errors. In the case of the middle image, a different branch of the correct vessel has been selected. For the rightmost image the main axis of the model is not correct causing an error for almost all vessel landmarks although the vascular arch as such has been successfully located.

OD localization most vessel points have moved along the arch. These errors point to a weakness in our evaluation method, for a more fair evaluation the entire venous vascular arch should be indicated in the reference standard.

The results of fitting the mean shape already seem quite good. This is caused by the fact that these images have been acquired using a protocol. The protocol ensures that the macula is located approximately in the same part of the image across the training and test set. This is evident in the results for the fovea detection which are good for most methods. If we would evaluate the vessel landmarks as suggested in the previous paragraph the difference in performance between the mean shape and the other methods would become larger.

The system as it is proposed here has several applications. An example would be to use it to automatically determine the severity of lesions related to eye diseases such as Diabetic Retinopathy. For lesions and other pathologies, their location on the retina is an important property. If lesions are located within the vascular arch the need for treatment becomes more immediate. Another application would be the automatic removal of the optic disc area before further processing by, for example, a computer aided diagnosis (CAD) system. This type of system can benefit from OD removal because of the a-typical appearance of the OD with respect to the rest of the retina.

The presented results could be accurate enough to be used as input for a CAD system. However, the proposed model is limited in the sense that the anatomical landmarks can only be found in macula centered retinal images. This suggests an extension to images

that are for instance OD centered or images which do not contain the fovea. Two ways to handle this extension would be to either built more models and use the separate models for the different cases or generalize the current model to enable it to segment other types of images as well.

References

[1] S. Arya, D.M. Mount, N.S. Netanyahu, R. Silverman, and A.Y. Wu. An Optimal Algorithm for Approximate Nearest Neighbor Searching in Fixed Dimensions. *Journal of the ACM*, 45(6):891–923, 1998.

[2] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.

[3] T. F. Cootes and Taylor. On representing edge structure for model matching. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1114–1120, 2001.

[4] T. F. Cootes, C. J. Taylor, D. Cooper, and J. Graham. Active shape models – their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[5] A. Corana, M. Marchesi, C. Martini, and S. Ridella. Minimizing multimodal functions of continuous variables with the "simulated annealing" algorithm. *ACM Transactions on Mathematical Software*, 13(3):262–280, 1987.

[6] D. Eberly. Skeletonization of 2d binary images, 2001. http://www.magic-software.com/Documentation/Skeletons.pdf.

[7] M. Foracchia, E. Grisan, and A. Ruggeri. Detection of optic disk in retinal images by means of a geometrical model of vessel structure. *IEEE Transactions on Medical Imaging*, 23(10):1189–1195, 2004.

[8] S. Kirkpatrick, C.D. Gelatt, and M.P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[9] H. Li and O. Chutatape. Automated feature extraction in color retinal images by a model based approach. *IEEE Transactions on Biomedical Engineering*, 51(2):246–254, 2004.

[10] J.A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, pages 308–313, 1965.

[11] M. Niemeijer, J.J. Staal, B. van Ginneken, M. Loog, and M.D. Abramoff. Comparative study of retinal vessel segmentation methods on a new publicly available database. In J. Michael Fitzpatrick and M. Sonka, editors, *SPIE Medical Imaging*, volume 5370, pages 648–656. SPIE, SPIE, 2004.

[12] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in C*. Cambridge University Press, Cambridge, 2nd edition, 1999.

[13] D. Ruprecht and H. Müller. Image warping with scattered data interpolation. *IEEE Computer Graphics and Applications*, 15(2):37–43, 1995.

[14] S. Sclaroff and J. Isidoro. Active blobs: Region-based, deformable appearance models. *IEEE Computer Graphics and Applications*, 89(2-3):197–225, 2003.