Sparse MRF Appearance Models for Fast Anatomical Structure Localisation*

René Donner[12], Branislav Mičušík[2],
Georg Langs[31], Horst Bischof[1]
[1]Institute for Computer Graphics and Vision, Graz University of Technology, Austria, bischof@icg.tugraz.at
[2]Pattern Recognition and Image Processing Group, Vienna University of Technology, Austria
{donner,micusik}@prip.tuwien.ac.at
[3]GALEN Group, Laboratoire de Mathématiques Appliquées aux Systèmes, Ecole Centrale de Paris, France, georg.langs@ecp.fr

Abstract

Image segmentation methods like active shape models, active appearance models or snakes require an initialisation that guarantees a considerable overlap with the object to be segmented. In this paper we present an approach that localises anatomical structures in a global manner by means of *Markov Random Fields (MRF)*. It does not need initialisation, but finds the most plausible match of the query structure in the image. It provides for precise, reliable and fast detection of the structure and can serve as initialisation for more detailed segmentation steps.

*Sparse MRF Appearance Models (*SAMs*)* encode a priori information about the geometric configurations of interest points, local features at these points and local features along the edges of adjacent points. This information is used to formulate a Markov Random Field and the mapping of the modeled object (e.g. a sequence of vertebrae) to the query image interest points is performed by the MAX-SUM algorithm.

The local image information is captured by novel symmetry-based interest points and local descriptors derived from *Gradient Vector Flow*. Experimental results are reported for two data-sets showing the applicability to complex medical data.

1 Introduction

The reliable and fast detection and segmentation of anatomical structures is a crucial issue in medical image analysis. It has been tackled by a number of powerful approaches,

*We would like to thank Philipp Peloschek, MD and Klaus Friedrich, MD of the Department of Radiology, Medical University of Vienna, Austria, for supplying the medical images. This research has been supported by the Austrian Science Fund (FWF) under grant P17083-N04 (AAMIR), as well as the European Union Network of Excellence FP6-507752 (MUSCLE) and the Region Île-de-France.

among them active shape models [3], active appearance models [4, 5], active feature models [12], graph-cuts [2] and snakes [7].

These approaches have been successfully employed to segment structures in cardiac MRIs [16] or for registration in functional heart imaging [19]. In [17] vertebrae in the spine were delineated, and in [20] active shape models were utilised for bone densiometry.

All approaches rely on a reasonable initialisation of the iterative active appearance model or active shape model search: ASMs and AAMs need to be placed with considerable overlap with the object of interest. Graph-cuts need a set of manually annotated seed points placed within and outside of the object, and while snakes need spatial constraints, to ensure the delineation of the correct object. Usually the initialization is either done manually or by application specific approaches.

Several approaches to a detect coarse initialization positions rely on pair-wise point matching using local descriptors like SIFT [13], shape context [1] or PCA-SIFT [8], and typically rely on a robust method like RANSAC [6] to deal with ambiguous structures. They match interest points between a source (i.e. example) image and the until now unseen target image. These approaches have several drawbacks: (1) For complex non-rigid transformations between source and target image a large number of correct interest points matches is required, which increases computation time considerably for the robust matching. (2) Information about the spacial relation of adjacent descriptors is difficult to incorporate into the matching process.

In this paper we propose a deterministic method based on Markov Random Fields (MRF) that incorporates both interest point positions and local features to perform the detection of landmark configurations from a single example. The detection is performed in a fast manner by the MAX-SUM algorithm [21]. The approach uses all interest point features and positions and finds a solution which minimizes the combined costs of non-rigid deformations and local descriptor feature differences. Arbitrary interest points and local descriptors can be used. We report results for interest points based on local symmetry and a complementary local descriptor derived from gradient vector flow [22].

Local symmetry detectors were investigated in [15, 10], but they are either computationally expensive or use radial symmetry detection with predefined radii. Recently [14] proposed an approach to detect symmetry in the constellation of interest points detected by existing point detection methods.

The paper is structured as follows: In Sec. 2 we explain the interest point detector and local descriptor. Sec.3 details Markov Random Fields and in Sec. 4 the mapping of the source- to the target points by MRFs are explained in detail. In Sec. 5 we present the experimental evaluation of our approach, followed by a conclusion and an outlook in Sec. 6.

2 Symmetry based interest points and descriptors

Many structures of interest to medical experts, like bones, veins and many anatomical structures or their parts exhibit a shape with a high degree of symmetry w.r.t. one or more axes. This property of (local) symmetry is well preserved even when dealing with 2D slices of 3D data sets like MRIs, as the cross sections of these body parts will appear as round or elongated structures. Even regions of interest that do not exhibit this property can

be localized by observing their neighborhood, e.g. an initialization for e.g. meniscoids can be provided by correctly localizing the discs and vertebrae of the spine.

2.1 Interest Points from Local Symmetry

Popular interest point detectors which are often used in conjunction with SIFT are the Harris corner detector and the Difference of Gaussians (DoG) approach, neither of which possess an affinity to local symmetry. A comparison of the interest points detected by DoG and interest points derived from local symmetry is shown in Fig. 1 (a,b).

To detect points of high local symmetry we use the gradient vector flow (GVF) field, which was originally proposed in [22] to increase the capture range of active contours. Its strengths include the ability to detect even weak structures while being robust to high amounts of noise in the image. The GVF can be computed either from a binary edge map or directly from the gray level image $\mathbf{I}$. We compute the GVF of an image as $\mathbf{G} = u + i * v = GVF(\mathbf{I})$, yielding the complex matrix $\mathbf{G}$ used for all subsequent computations. The resulting field $\mathbf{G}$ is depicted in Fig. 2 for a synthetic example and a section of a hand radiograph, overlaid over the image $\mathbf{I}$.

The field magnitude $|\mathbf{G}|$ is largest in areas of high image gradient, and the start- and endpoints of the field lines of $\mathbf{G}$ are located at symmetry maxima. E.g. in the case of a symmetrical structure formed by a homogeneous region surrounded by a different gray level value the field will point away form or towards the local symmetry center of the structure, as shown in Fig. 2 (a,b). The symmetry interest points are thus defined as the local minima of $|\mathbf{G}|$. In contrast to techniques based on estimating the radial symmetry using a sliding window approach this will yield a sparse distribution of interest points even in large homogeneous regions.

After detecting the interest points the orientation $\alpha_i \in [0, 2\pi[$ of the local region surrounding the interest point can be estimated. Around each interest point rays $\mathbf{g}_\alpha^r$ at the 360 angles $\alpha \in [0, \ldots, 2\pi[$ at radii $r \in \{2, \ldots, 8\}$ are sampled from $|\mathbf{G}|$ using bilinear interpolation. The interest point i is then assigned the angle α_i which minimises

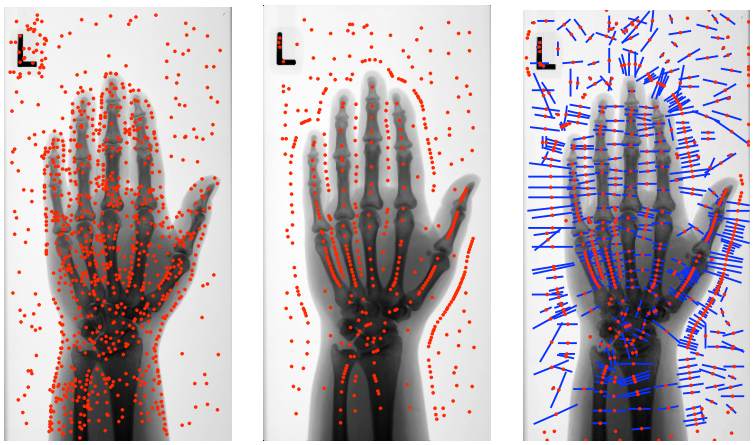
$$\alpha_i = \underset{\alpha \in [0, 2\pi[}{\operatorname{argmin}} \sum_r \mathbf{g}_\alpha^r. \tag{1}$$

The scale s_i of the region around the interest point is estimated by the mean distance of the interest point i to the two closest local maxima of $|\mathbf{G}|$ in the directions of α_i and $\alpha_i + \pi$. Examples for the resulting estimates for orientation and scale are shown in Fig 1 (c). If the scale varies only within a limited range as for the medical images examined in this paper the scale can remain fixed.

2.2 Local Descriptors from Gradient Vector Flow Fields

A measure is needed to specify the similarity of the local regions around the symmetry interest points and edges. Several local descriptors have been proposed in recent years, including SIFT [13] and Shape Context [1]. While most of these approaches yield descriptors suitable for building the MRF, they would require additional computations. In contrast, we can directly use $\mathbf{G}$ to describe local context.

In [8] normalized patches of the image gradient are used, extracted according to the interest points' orientation and scale as local descriptor are. Similarly, we extract patches



(a) (b) (c)

Figure 1: Comparison of the (a) interest points found by Difference of Gaussians (DoG) and (b) the symmetry points found as minima of GVF magnitude. Note how the symmetry points pick up the structures which are of interest to medical experts, greatly facilitating the correct localization of these structures. (c) depicts the scale and orientation estimates obtained around the symmetry points.

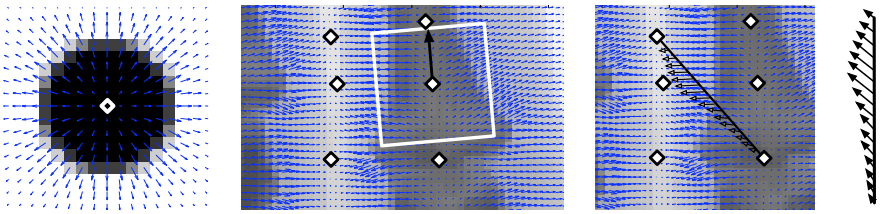
of $\mathbf{G}$ around the symmetry interest points, according to scale s_i and orientation α_i, as depicted in Fig. 2. They are re-sampled to a 10×10 grid and the vector field's orientations are stored relative to α_i to form the actual local descriptor. This encodes the information about the image gradients within and around the patch in a rotation-invariant way. Because of the GVF's smooth structure, Euclidean distance can be used used to compute the distance between two descriptors. This eliminates the need for complex histogram construction as performed by SIFT for example, while still retaining a feature vector of low dimensionality.

As the orientation of the local interest point is usually only stable up to $\pm\pi$, the actual distance between two local descriptors $\mathbf{D}_1$ and $\mathbf{D}_2$ is computed as $min(\|\text{abs}(\mathbf{D}_1 - \mathbf{D}_2)\|, \|\text{abs}(\mathbf{D}_1 - \mathbf{D}_2^*)\|)$, where $\mathbf{D}_2^*$ denotes the descriptor 2 rotated by π.

Local edge descriptors In addition to the local descriptors around interest points the appearance along the models' edges forms an important part of sparse appearance models.

Again the GVF $\mathbf{G}$ is used to extract the relevant information. Given 2 interest points in the image, $\mathbf{G}$ is sampled at equidistant points along the edge. The sampled field is then stored relative to the edge's orientation, forming a complex vector $\mathbf{e}$ as displayed in Fig. 2 (d). For the experiments in this paper, 40 points were sampled per edge.

By describing an image using the GVF-based local descriptors around interest points and along edges, the essential information about the structure of the anatomical object is captures in a sparse fashion. Sec. 4 describes how the descriptors from several training images are combined to form a sparse appearance model.



(a) (b) (c) (d)

Figure 2: (a) Examples of GVF with the detected symmetry interest points (diamonds). (b) Descriptor extraction from the GVF field. Around each symmetry point patches are extracted from the vector field according to their scale and orientation. The patch is then resampled to a 10×10 grid, relative to the interest points' orientation, to form the actual descriptor. The image is displayed for better visualization, the symmetry points are marked as diamonds. (c) Schematic edge descriptor of the edge between two points, formed by sampling the GVF at equidistant points along the edge. (d) Resulting edge descriptor, relative to the edge's orientation.

3 Markov Random Fields and the MAX-SUM problem

The Markov Random Fields considered in this paper represent graphs where each of the M nodes, called objects, has N fields, or labels, with associated qualities. The labels of two adjacent nodes are fully connected by N^2 edges, again with a weight to encode quality. Which objects are adjacent is encoded in an additional graph $\mathscr{A}$ with A edges. This basic structure is depicted in Fig. 3 (a). There are 4 objects with 3 labels each, with $N^2 = 9$ edges between the adjacent objects, A is 5.

Of interest is now to select one label for each object, so that the sum of label and edge qualities of the resulting sub-graph becomes maximal, illustrated as thick lines. The MAX-SUM solver can be used to tackle this problem. The MAX-SUM (labeling) problem of the second order is defined as maximizing a sum of bivariate functions of discrete variables. The solution of a MAX-SUM problem corresponds to finding a configuration of a Gibbs distribution with maximal probability. It is equivalent to finding a maximum posterior (MAP) configuration of an MRF with discrete variables [21].

Let the $M \times N$-matrix $\mathbf{C}$ represent the label qualities for each of the objects, and the $A \times N^2$-matrix $\mathbf{E}$ represent the edge qualities between the pairs of labels.

The total quality of the label selection $\mathbf{S} = \{n_1, \ldots, n_M\}$ with $n_i \in \{1, \ldots, N\}$ is then defined as

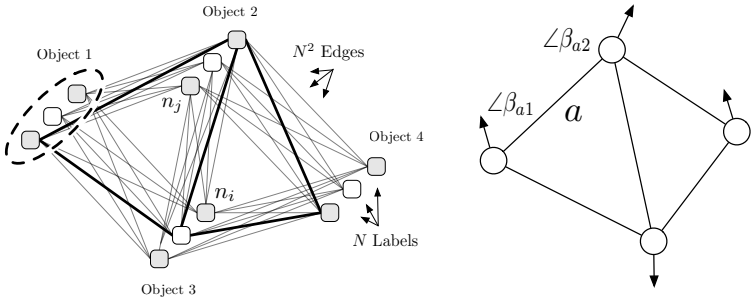
$$C(\mathbf{S}) = \sum_{m=1\ldots M} \mathbf{C}(m, \mathbf{S}(m)) + \sum_{a=1\ldots A} \mathbf{E}(a, \beta(E, \mathbf{S}, a)), \tag{2}$$

where $\beta(\mathbf{E}, \mathbf{S}, a)$ denotes the column representing the quality of the edge between the labels chosen to represent the edge $\mathscr{A}(a)$.

Solving the MAX-SUM problem means finding the set of optimal labels

$$\mathbf{S}^* = \underset{\mathbf{S}}{\operatorname{argmax}}\, C(\mathbf{S}). \tag{3}$$

Recently, a very efficient algorithm for solving this problem through linear programming relaxation and its Lagrangian dual, originally proposed by Schlesinger in 1976 [18],



(a) (b)

Figure 3: (a) The MRF graph consists of M objects with N labels each. Qualities are assigned to both labels and edges. Finding the solution to a MAX-SUM problem means selecting a label for each object, such that the sum of qualities of the selected labels and the edges connecting them is maximized. (b) Illustration of how the relative angles between an edge and the orientations of its adjacent vertices is computed.

has been presented [21].

The MAX-SUM solver permits several labels to be defined while still keeping the processing time within reasonable bounds. There are other attempts to solve the labeling problem for MRF using, e.g., second order cone programming [11], sequential tree-reweighted max-product message passing [9] or belief propagation methods [23]. However, neither of the algorithms, nor the MAX-SUM approach, solve the problem of a multi-label MRF exactly, as it is NP-hard. If the graph is a tree the global optimum of Eq. (3) is guaranteed [9], in the case of a non-tree graph MAX-SUM takes various approximations into account to reach a possibly optimal solution.

4 Sparse Appearance Model Matching

This sections describes how the sparse appearance model is constructed from training data. This model is then used to specify the Markov Random Field for a target image.

Building a Sparse Appearance Model Sparse appearance models extract information from images using local descriptors around interest points and along the edges between these points. No PCA based model is used to avoid the need for a large number of training samples and the global character of PCA-based models. The shape model is based on a Delaunay triangulation of the model points, and statistical models of the edges' lengths, relative angles and local descriptors are recorded. This yields a locally deformable rotation invariant model. The interest points and local point/edge descriptors are based on local symmetry and GVF as described in Sec. 2.

For each of the n model images, a subset of M interest points is manually selected to describe the anatomical structure to be found. One of the model images is used to define the graph structure using a Delaunay triangulation of its M model points. The

resulting adjacencies of model points yield the set $\mathscr{A}$ of index-tuples describing the edges. Examples of the generated model are shown in Fig. 4 (a,b).

The M selected model points represent the objects of the MRF graph, while the N target interest points correspond to the labels. A solution $\mathbf{S}$ thus represents a mapping of the model interest points to a subset of the target interest points.

We now need to build the a priori statistical models from the n training samples for the M model points and the A edges between these model points. First the orientations of the model points are normalized. As the n training orientations for a model point m are only stable up to $\pm\pi$, π is added to a subset of them such that the circular variance of the n orientations of model point m is minimised.

As there are generally too few training samples to estimate the parameters of a multivariate Gaussian in the space of the local descriptors, only the mean of the n local descriptors for each model point m is used, yielding descriptors $\overline{\mathbf{D}}_m$.

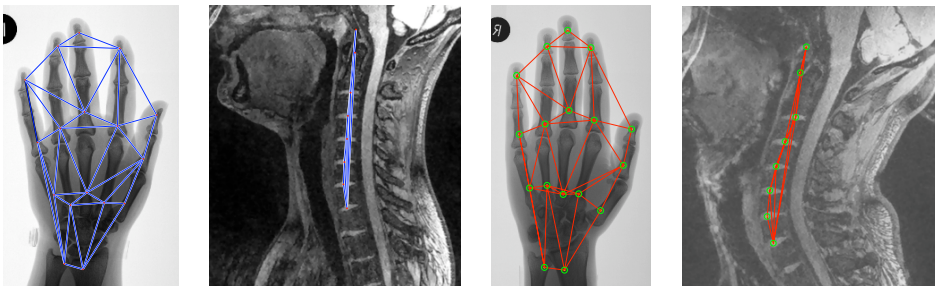
For each edge a of the A model edges, the mean length $\bar{l}_a$ and the standard deviation l_a^σ is computed. Similarly, from the n angles β_{a1} and β_{a2} between the edge and the orientations of its vertices the mean angles and standard deviations $\overline{\beta}_{a1}$,$\overline{\beta}_{a2}$ and β_{a1}^σ,β_{a2}^σ, computed using circular statistics, are stored (see Fig. 3 (b)). The third edge property which is modelled is the local descriptor (see Fig. 2). Similarly to the point descriptors, the mean descriptor $\bar{\mathbf{e}}_a$ is computed for each model edge.

Constructing the MRF Given a sparse appearance model and a target image, the Markov Random field is used to model the confidences that a model point or edge should be matched to a certain interest point or edge in the target image. As we are solving a maximization problem, all confidences or qualities are in the interval $[-\infty, 0]$. The descriptor distances are normalized to having a maximum of 0 and a median of -1, while the length and angle confidences are $\in [-1, 0]$.

The quality of a (model point, target point)-match equals the negative distance between the local target descriptor and the model point descriptor $\overline{\mathbf{D}}_m$. All mutual distances between model and potential target correspondences are computed, resulting in the $M \times N$-matrix $\mathbf{C}$.

The qualities of the AN^2 edges in the model are stored in $\mathbf{E}$. The quality of an edge e between two labels n_i, n_j in $\mathbf{E}$ is computed by comparing its length l_e and relative angles β_{e1}, β_{e2} with the corresponding Gaussian distributions of the model edge ($\bar{l}_a, l_a^\sigma$, $\overline{\beta}_{a1}$,$\overline{\beta}_{a2}$, β_{a1}^σ,β_{a2}^σ). Identity with the mean yields a confidence of 0, the minimum confidence is -1. The confidence for the edge's appearance equals the negative distance between the edge descriptor and the model edge descriptor $\bar{\mathbf{e}}_a$. The overall confidence of edge e representing the model edge a is finally set to the minimum of the confidences for length, angles and descriptor, thus effectively filtering out unlikely candidates.

It can occur that no interest point is detected in one location of the medical structure in the target image where the model would expect one. It is thus important to include the possibility of omitting a model point. This is achieved by adding one artificial target interest point (dummy point), yielding $\mathbf{Cd}$ and $\mathbf{Ed}$ of sizes $M \times N + 1$ and $A \times (N+1)^2$, respectively. The last column of $\mathbf{Cd}$ is set to the mean of $\mathbf{C}$ multiplied by a factor f controlling how costly it should be to omit a model point. Similarly, the edges of $\mathbf{Ed}$ involving the dummy point are set to f times the mean of $\mathbf{E}$.



(a) (b) (c) (d)

Figure 4: (a,b) Model graph $\mathscr{A}$ automatically generated from the M selected interest points (landmarks) depicted for two of the training images. (c,d) The results of the model matching for two test images.

The MAX-SUM solver is then applied on **Cd**, **Ed**, yielding the set $\mathbf{S} = \{n_1, \ldots, n_M\}$ of optimal labels for each model node, maximizing the quality C in Eq. 3.

5 Experiments

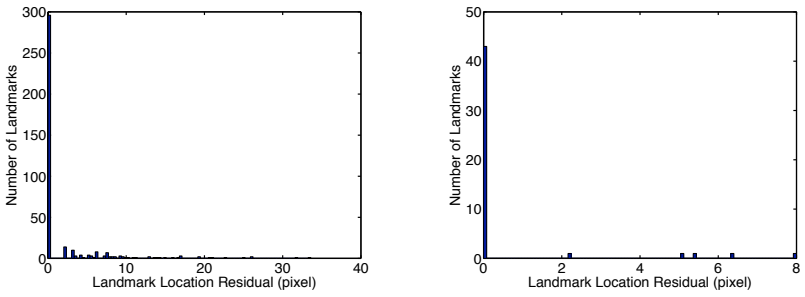
The approach was evaluated[1] on 2 data sets (Fig. 4). **1.** For a set of 25 hand radiographs (300×450 pixels) 17 landmarks in each image were manually annotated. **2.** On 8 spine MRIs (280×320 pixels) manual annotations of the centers of 6 inter-vertebral discs were used for validation, plus 2 landmarks to disambiguate the matching. The error between found landmarks and ground truth landmarks was recorded, where only the points of medical interest (only the 6 spine landmarks which correspond to vertebral discs) where considered. The typical number of detected interest points was between 400 and 600, the model graphs contained 17 and 8 nodes, respectively. In Fig. 4 (a,b) the model graphs are depicted on two of the training images. In Fig. 4 (c,d) matching results are depicted: the red lines represent the model graph matched to the target image, while the green circles are the positions of the found landmarks.

Quantitative analysis was performed by a leave-one-out procedure i.e a single image was chosen as target image and the model graph was built from the the remaining 24 or 7 images respectively. The mean/median error for matches is 2.79 / 0.0 pixels for hand data and 0.56 / 0.0 pixels for the spine data, reflecting the excellent matching accuracy. The error histograms for both sets are depicted in Fig. 5. Typical run times for solving the MRF are in the order of few seconds.

6 Conclusion and Outlook

We present a framework for the fast and accurate localisation of anatomical structures. Configurations of symmetry interest points and local descriptors derived from Gradient Vector Flow are represented by graphs and Markov Random Fields. The matching is

[1] The implementation used in this evaluation is available at `http://www.aamir.at/bmvc07/`



(a) (b)

Figure 5: Result histograms for the pixel distances of result landmarks to ground truth landmarks for (a) the hand radiograph data set and (b) the spine MRI data set. Note the high quality of the model matching, with most of the landmarks being matched perfectly.

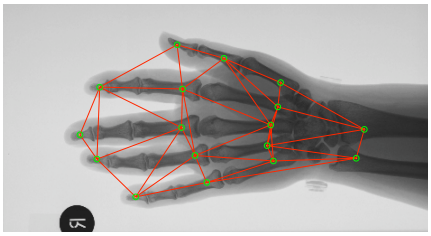


Figure 6: Example of the rotation invariance of Sparse Appearance Models: The model was trained on upright hand radiographs. As only relative angles are modeled, the hand is successfully detected in the rotated image.

performed by the MAX-SUM algorithm. The approach integrates local descriptor similarities and deformation constraints in a single optimization step. Results indicate that the method provides the localization accuracy necessary for the initialization of subsequent segmentation algorithms. Future research will focus on improvements to allow for the application to segmentation tasks as well as the extension to 3-dimensional data sets.

References

[1] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE PAMI*, 24(4):509–522, 2002.

[2] Y. Boykov and M.-P. Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. In *Proc. ICCV*, pages 105–112, 2001.

[3] T. Cootes. Active shape models - ‘smart snakes’. In *Proc. BMVC*, 1992.

[4] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Trans. PAMI*, 23(6):681–685, 2001.

[5] R. Donner, M. Reiter, G. Langs, P. Peloschek, and H. Bischof. Fast active appearance model search using canonical correlation analysis. *IEEE TPAMI*, 28(10):1690 – 1694, October 2006.

[6] M. A. Fischler and R. C. Bolles. A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. of the ACM*, 24, 1981.

[7] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal on Computer Vision*, 1:321–331, 1988.

[8] Y. Ke and R. Sukthankar. PCA-SIFT: A more distinctive representation for local image descriptors. In *CVPR (2)*, pages 506–513, 2004.

[9] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *PAMI*, 28(10):1568–1583, 2006.

[10] P. Kovesi. Symmetry and asymmetry from local phase. In *Proceedings of the Tenth Australian Joint Conference on Artificial Intelligence*, pages 185–190, 1997.

[11] M. P. Kumar, P. H. S. Torr, and A. Zisserman. Solving Markov random fields using second order cone programming. In *Proc. CVPR*, pages I: 1045–1052, 2006.

[12] G. Langs, P. Peloschek, R. Donner, M. Reiter, and H. Bischof. Active feature models. In *Proc. ICPR*, pages 417–420, 2006.

[13] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 2004.

[14] G. Loy and J.-O. Eklundh. Detecting symmetry and symmetric constellations of features. In *Proceedings of ECCV '06*, 2006.

[15] G. Loy and A. Zelinsky. Fast radial symmetry for detecting points of interest. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(8):959–973, 2003.

[16] S. C. Mitchell, J. G. Bosch, B. P. F. Lelieveldt, R. J. van der Geest, J. H. C. Reiber, and M. Sonka. 3-d active appearance models: Segmentation of cardiac MR and ultrasound images. *IEEE TMI*, 21(9):1167–1178, 2002.

[17] M. Roberts, T. F. Cootes, and J. E. Adams. Vertebral morphometry - semiautomatic determination of detailed shape from dual-energy x-ray absorptiometry images using active appearance models. *Investigative Radiology*, 41(12):849–459, 2006.

[18] M. Schlesinger. Sintaksicheskiy analiz dvumernykh zritelnikh signalov v usloviyakh pomekh (syntactic analysis of two-dimensional visual signals in noisy conditions). *Kibernetika*, (4):113–130, 1976. In Russian.

[19] M. B. Stegmann, H. Ólafsdóttir, and H. B. W. Larsson. Unsupervised motion-compensation of multi-slice cardiac perfusion MRI. *Medical Image Analysis*, 9(4):394–410, aug 2005.

[20] H. Thodberg and A. Rosholm. Application of the active shape model in a commercial medicals device for bone densitometry. In *BMVC*, pages 43–52, 2001.

[21] T. Werner. A linear programming approach to Max-sum problem: A review. *IEEE Trans. on Pattern Recognition and Machine Intelligence*, 29(7), 2007.

[22] C. Xu and J. L. Prince. Snakes, shapes, and gradient vector flow. *IEEE Trans. on Image Proc.*, 7(3), March 1998.

[23] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Constructing free-energy approximations and generalized belief propagation algorithms. *IEEE Transactions on Information Theory*, 51(7):2282–2312, 2005.