

Object Localization Based on Markov Random Fields and Symmetry Interest Points^{*}

René Donner^{1,2}, Branislav Micusik², Georg Langs^{1,3}, Lech Szumilas²,
Philipp Peloschek⁴, Klaus Friedrich⁴, and Horst Bischof²

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,lech}@prip.tuwien.ac.at`

³ GALEN Group, Laboratoire de Mathématiques Appliquées aux Systèmes,
Ecole Centrale de Paris, France

`georg.langs@ecp.fr`

⁴ Department of Radiology, Medical University of Vienna, Austria
`{philipp.peloschek,klaus.friedrich}@meduniwien.ac.at`

Abstract. We present an approach to detect anatomical structures by configurations of interest points, from a single example image. The representation of the configuration is based on Markov Random Fields, and the detection is performed in a single iteration by the MAX-SUM algorithm. Instead of sequentially matching pairs of interest points, the method takes the entire set of points, their local descriptors and the spatial configuration into account to find an optimal mapping of modeled object to target image. The image information is captured by 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, among them are active shape models [3], active appearance models [4], and graph-cuts [2]. They have successfully been employed to segment structures in cardiac MRIs [13] or for registration in functional heart imaging [15]. These methods need to be initialized: ASMs and AAMs need to be placed with considerable overlap with the object of interest. Graph-cuts need manually annotated seed points placed within and outside of the object. This initialization is either done manually or by application specific approaches.

^{*} This research has been supported by the Austrian Science Fund (FWF) under grants P17083-N04 (AAMIR) and P17189-N04 (SESAME), as well as the European Union Network of Excellence FP6-507752 (MUSCLE) and the Region Île-de-France.

An approach to a detection of such initialization positions is to use local descriptors like SIFT [10], shape context [1] or PCA-SIFT [6]. They match interest points between a source (i.e. example) image and the until now unseen target image, and typically rely on a robust estimation method like RANSAC [5]. These approaches have several drawbacks: For complex non-rigid transformations between source and target image a large number of correct interest points matches is required to correctly estimate the unknowns of the transformation, which considerably increases computation time for the robust matching. Information about the spatial 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 single iteration by the MAX-SUM algorithm [16]. 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 [17].

Local symmetry detectors were investigated in [8,12], but they are either computationally expensive or use radial symmetry detectors of predefined radii. Recently [11] 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. In Sec.4 the mapping of the source- to the target points by MRFs will be 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. an axis. 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 them possessing an affinity to local symmetry. A comparison of the interest points detected by DoG and interest points derived from local symmetry

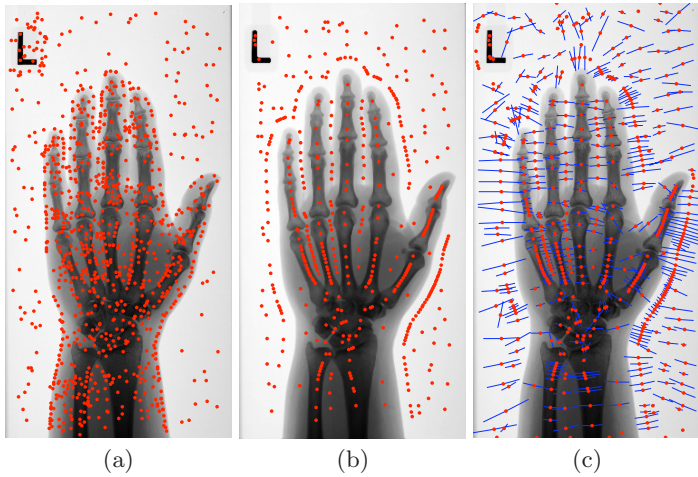


Fig. 1. Comparison of the (a) interest points found by difference of Gaussians (DoG) and (b) the symmetry points found as minima of GVF magnitude. Note how the symmetry points pick up the structures which are of interest to medical experts, greatly facilitating the correct localization of these structures.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

is shown in Fig. 1 (a,b). To detect points of high local symmetry we use the gradient vector flow (GVF) field originally proposed in [17] 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 when used for symmetry detection. To further reduce the influence of noise the image can be median-filtered prior to computing GVF. 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 synthetic examples 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 from 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 $|\mathbf{G}|$.

After detecting the interest points the orientation $b_i \in [0, \pi]$ of the local region surrounding the interest point can be estimated. It is computed as $b_i = \angle \mathbf{G}(x_i + \Delta x_i, y_i + \Delta y_i)$, which is the orientation of \mathbf{G} at a pixel in a local $r \times r$ -pixel neighborhood satisfying

$$(\Delta x_i, \Delta y_i) = \underset{\substack{\Delta y_i \in \{-r/2, \dots, r/2\} \\ \Delta x_i \in \{0, \dots, r/2\}}} {\operatorname{argmin}} \quad |(\angle \mathbf{G}(x_i + \Delta x_i, y_i + \Delta y_i) - \angle \mathbf{G}(x_i - \Delta x_i, y_i - \Delta y_i))|. \tag{1}$$

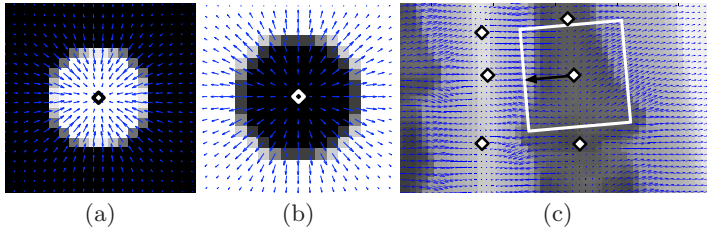


Fig. 2. (a,b) Examples of GVF with the detected symmetry interest points (diamonds). (c) Descriptor extraction from the GVF field. Around each symmetry point patches are extracted from the vector field according to their scale and orientation. The patch is then resampled to a 10×10 grid to form the actual descriptor. The image is displayed for better visualization, the symmetry points are marked as circles.

The scale s_i of the region around the interest point is estimated by the mean distance from (x_i, y_i) to the two closest local maxima of $|\mathbf{G}|$ in the direction of $b_i \pm \pi$. Examples for the resulting estimates for orientation and scale are shown in Fig 1 (c).

2.2 Descriptors from Gradient Vector Flow Fields

A measure is needed to specify the similarity of the local regions around the symmetry interest points. Several local descriptors have been proposed in recent years, including SIFT [10] 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.

[6] use normalized patches of the image gradient according to the interest points' orientation and scale as local descriptor. Similarly, we extract patches of \mathbf{G} around the symmetry interest points, according to scale s_i and orientation b_i . They are re-sampled to a 10×10 grid, as depicted in Fig. 2, to form the actual local descriptor. This encodes the information about the image gradients within and around the patch. Because of the GVF's smooth structure, Euclidean distance can be 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 only uniquely defined 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 .

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,

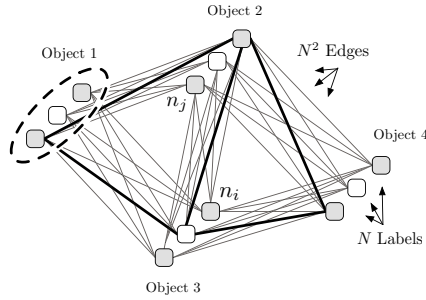


Fig. 3. The MRF graph consists of M objects with N labels each. Qualities are assigned to both labels and edges. Finding the solution to a MAX-SUM problem means selecting a label for each object, such that the sum of qualities of the selected labels and the edges connecting them is maximized.

again with a weight to encode quality. Which objects are adjacent is encoded in an additional graph \mathcal{A} with a edges. This basic structure is depicted in Fig. 3. 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 [16].

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 $\mathcal{S} = \{n_1, \dots, n_M\}$ with $n_i \in \{1, \dots, N\}$ is then defined as

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

where $\beta(\mathbf{E}, \mathcal{S}, \alpha)$ denotes the column representing the quality of the edge between the labels chosen to represent the edge $\mathcal{A}(\alpha)$. Solving the MAX-SUM problem means finding the set of optimal labels

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

Recently, a very efficient algorithm for solving this problem through linear programming relaxation and its Lagrangian dual, originally proposed by Schlesinger in 1976 [14], has been presented [16]. 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 [9], sequential tree-reweighted max-product message passing [7] or belief propagation methods [18]. 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 [7], in the case of a non-tree graph MAX-SUM takes various approximations into account to reach a possibly optimal solution.

4 Localization of Anatomical Structures

For a model image, a subset of interest points is manually selected to describe the medical object to be found. The Delaunay triangulation of these M model points yields the set \mathcal{A} of index-tuples describing the edges. An example of the generated model is shown in Fig. 4 (a,b).

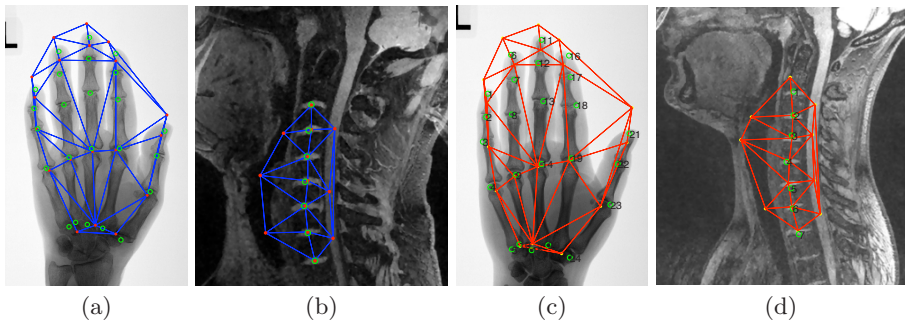


Fig. 4. (a,b) Model graph \mathcal{A} automatically generated from the symmetry points selected on the model image. The additionally placed landmarks (circles) are not part of the model and are used only for visualization. (c,d) show the graphs matched to test images, including the landmarks propagated according to the correspondences found by the matched graph.

The M selected model points represent the objects of the MRF graph, while the indices of the N target interest points correspond to the labels. A solution \mathcal{S} thus represents a mapping of the model interest points to a subset of the target interest points, assigning one target interest point to each model point.

The quality of a (model point, target point)-match equals the negative distance between their local descriptors (as we are solving a maximization problem). 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 between two labels n_i, n_j in \mathbf{E} is computed by comparing its length and angle with the edge between the corresponding objects (model nodes). As the medical structures under investigation can be assumed to be of similar scale, the edge quality e is set to

$$e(\alpha, n_i, n_j) = -(|\text{length}\mathcal{A}(\alpha) - \text{length}(n_y, n_z)| + \gamma(|\angle\mathcal{A}(\alpha) - \angle(n_i, n_j)|)), \quad (4)$$

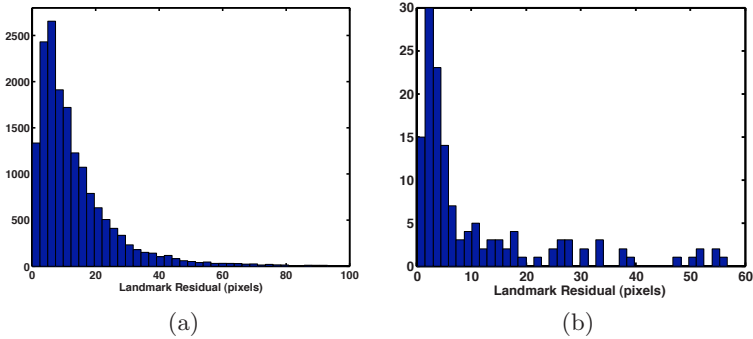


Fig. 5. Result histograms of the distances of propagated landmarks to standard of reference landmark positions for (a) the hand and (b) the spine data set

where $\text{length}(h, k)$ represents the pixel distance between interest points k and h , $\angle(h, k)$ is the orientation of the edge and γ is a normalization factor to compensate for the different scale of angles and lengths.

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} . The MAX-SUM solver is then applied on \mathbf{Cd} , \mathbf{Ed} , yielding the set $\mathbf{S} = \{n_1, \dots, n_M\}$ of optimal labels for each model node, maximizing the quality C in Eq. 3. The presented method thus in effect performs a non-rigid registration of the partial model image to the test image.

As the interest points are not necessarily at the locations medical experts are interested in, additional landmark points are manually set in the model image. They are not used for computing the match, but only for result visualization and evaluation.

5 Experiments

The approach was evaluated on 2 data sets (Fig. 4). **1.** For a set of 30 hand radiographs (300×450 pixels) standard of reference annotations (landmarks) for 24 joints in each image were available. **2.** On 5 spine MRIs (280×320 pixels) manual annotations of 7 inter-vertebral discs were used. To evaluate the matching accuracy the landmarks were propagated according to the match and the pixel error between propagated and correct landmarks was recorded. Piecewise affine transformation of the Delaunay triangulation of the selected source symmetry points is used to propagate the source landmarks to the target image. The typical number of detected interest points was between 400 and 600, the

model graphs contained 10 to 25 nodes. In Fig. 4 (a,b) the source model graphs for two examples are depicted. The model graph is depicted by blue lines, green circles are manual annotations used only for validation. In Fig. 4 (c,d) matching results are depicted: red lines represent the model graph matched to the target image, while green circles indicate the propagated landmarks. Quantitative analysis was performed by a leave-one-out procedure i.e a single image was chosen as source and the model graph was matched to the remaining 29 or 4 images. The mean/median error for matches is 14.2 / 9.7 pixels for hand data (a typical joint width is 25 pixels) and 10.85 / 4.8 pixels for the spine data. This is sufficient for most initialization purposes. The error histograms in Fig. 5 show the pixel distances for all propagated landmarks to the correct target landmark positions from all runs. Typical run times for solving the MRF for one source-target match are in the order of a few seconds.

6 Conclusion and Outlook

We present a framework for the matching of anatomical structures from a single example. Configurations of interest points are represented by graphs and Markov random fields, and the matching is performed in one iteration 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 using combined model graphs from several model images, and the extension to 3D data sets.

References

1. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. *IEEE PAMI* 24(4), 509–522 (2002)
2. Boykov, Y., Jolly, M.-P.: Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. In: *Proc. ICCV*, pp. 105–112 (2001)
3. Cootes, T.: Active shape models - ‘smart snakes’. In: *BMVC* (1992)
4. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. *IEEE Trans. PAMI* 23(6), 681–685 (2001)
5. Fischler, M.A., Bolles, R.C.: A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. of the ACM* 24 (1981)
6. Ke, Y., Sukthankar, R.: PCA-Sift: A more distinctive representation for local image descriptors. In: *CVPR* (2), pp. 506–513 (2004)
7. Kolmogorov, V.: Convergent tree-reweighted message passing for energy minimization. *PAMI* 28(10), 1568–1583 (2006)
8. Kovesi, P.: Symmetry and asymmetry from local phase. In: *Proceedings of the Tenth Australian Joint Conference on Artificial Intelligence*, pp. 185–190 (1997)
9. Kumar, M.P., Torr, P.H.S., Zisserman, A.: Solving Markov random fields using second order cone programming. In: *Proc. CVPR*, pp. I:1045–1052 (2006)
10. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *IJCV* (2004)

11. Loy, G., Eklundh, J.-O.: Detecting symmetry and symmetric constellations of features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) *ECCV 2006*. LNCS, vol. 3952, Springer, Heidelberg (2006)
12. Loy, G., Zelinsky, A.: Fast radial symmetry for detecting points of interest. *IEEE Trans. Pattern Anal. Mach. Intell.* 25(8), 959–973 (2003)
13. Mitchell, S.C., Bosch, J.G., Lelieveldt, B.P.F., van der Geest, R.J., Reiber, J.H.C., Sonka, M.: 3-d active appearance models: Segmentation of cardiac MR and ultrasound images. *IEEE TMI* 21(9), 1167–1178 (2002)
14. Schlesinger, M.: Sintaksicheskiy analiz dvumernykh zritelnykh signalov v usloviyakh pomekh (syntactic analysis of two-dimensional visual signals in noisy conditions). *Kibernetika* (4), 113–130 (1976) (in Russia)
15. Stegmann, M.B., Ólafsdóttir, H., Larsson, H.B.W.: Unsupervised motion-compensation of multi-slice cardiac perfusion MRI. *Medical Image Analysis* 9(4), 394–410 (2005)
16. Werner, T.: A linear programming approach to Max-sum problem: A review. Research Report CTU–CMP–2005–25, Czech Technical University (2005)
17. Xu, C., Prince, J.L.: Snakes, shapes, and gradient vector flow. *IEEE Trans. on Image Proc.* 7(3) (March 1998)
18. Yedidia, J.S., Freeman, W.T., Weiss, Y.: Constructing free-energy approximations and generalized belief propagation algorithms. *IEEE Transactions on Information Theory* 51(7), 2282–2312 (2005)