Color Symmetry for Interest Point Detection

Gunther Heidemann
Intelligent Systems Group, University of Stuttgart
Universitätstr. 38, D-70569 Stuttgart, Germany

Abstract

This paper presents an algorithm for interest point detection based on symmetry features. Interest points are used in computer vision to represent salient image areas which have an increased likelihood to be relevant for further processing. In contrast to most other works, the approach presented here uses color as a form feature on equal footing with intensity. Thus it is possible to detect symmetric structures even in the presence of low contrast and on inhomogeneous background. The application of interest points is described in the context of object recognition and image retrieval.

1 Introduction

The detection of interest points (IPs) is a key component in many image processing architectures, e.g. for object recognition, image retrieval, or active vision [17, 1]. The role of IPs in image processing is to direct the "attention" of the system to the areas which are most relevant for further processing. It is usually argued that restricting processing to a limited number of image areas is for the sake of computational efficiency, but it should be noted that the pre-selection of certain patterns is also part of the pattern recognition process itself.

While segmentation techniques (e.g. [11, 3]) partition an image into regions which are well-defined by their borders, an IP represents a "salient" image area A only by a single point. But A is not a precisely defined region in the sense that a border can be given such that the image is salient inside the border and not salient on the outside. The size and shape of the salient area A heavily depends on the IP-algorithm, for a discussion on this problem see [7]. This difference between segmentation and IP algorithms is due to their functional principles: Segmentation searches for homogeneity of features such as color or texture, whereas IP algorithms look for structures such as edges or corners.

So far, practically all IP-algorithms exploit gray values only. The reason for this restriction is that IPs are found from form features (e.g. [15]), and form features are widely believed to be detectable only from intensity variation. This assumption roots in early findings on biological vision, where shape is supposed to be perceived from intensity, whereas color information is added to shapes in later processing stages (e.g. [9]).

But experience as well as recent analysis of the second order statistics of color imagery has shown that color is a promising form feature [8], though spatio-chromaticity is still rarely used in technical vision. This paper presents an IP-detector which exploits the spatial distribution of color and gray values alike. It is based on the well known symmetry detector proposed by Reisfeld et al. [12]. But while the old approach has problems in detecting symmetries in the presence of low contrast, the new algorithm can exploit color differences.

Organization of the paper: In the following section 2, first the original gray value based approach will be described, then the extension to color and the choice of color spaces. Section 3 describes two complementary applications: For the recognition of pre-defined objects or parts of objects, IPs are used in a “one IP – one object” fashion (section 3.1), whereas the characterization of images from unknown domains for retrieval purposes requires multiple IPs (section 3.2). The final section 4 gives a short summary and outlook.

2 Color based IP detection

The techniques considered in this paper are context-free, i.e. independent of a particular image domain and not goal specific (i.e. specialized to localizing pre-defined objects). There are numerous approaches to context-free IP detection. Most are aimed at finding edges and / or corners, the most well known is the detector of Harris and Stephens [4]. By contrast, Reisfeld et al. have proposed an algorithm which yields IPs in the middle of symmetric regions. This algorithm will be addressed as GRAY-SYM. GRAY-SYM yields a *saliency map*, for an example see the upper left picture in Fig. 1, where bright spots indicate high symmetry in the original (upper right). However, GRAY-SYM often fails to detect symmetry in the presence of low contrast, as visible in the second upper right picture, where the markers are the IPs obtained as the maxima of the saliency map. The algorithm and the reasons for its failures are discussed in the following section 2.1. In section 2.2 the generalization to color (COL-SYM) will be outlined.

2.1 The GRAY-SYM algorithm

The GRAY-SYM algorithm used here as a basis for COL-SYM is almost identical to the algorithm proposed by Reisfeld et al. [12] with minor modifications for computational efficiency. GRAY-SYM calculates a continuous valued symmetry judgment for each point of the image. Let the image be given by its gray values $I(p)$ where p denotes a pixel at location (x, y). The gradient of $I(p)$ is denoted by $(I_x(p) = \frac{\partial I(p)}{\partial x}, I_y(p) = \frac{\partial I(p)}{\partial y})$, from which the gradient magnitude $G_I(p) = \sqrt{I_x(p)^2 + I_y(p)^2}$ and direction $\theta_I(p) = \arctan\,(I_y(p)/I_x(p))$ can be calculated.

A set $\Gamma(p)$ of index pairs (i, j) of pixel pairs (p_i, p_j) which surround the central pixel p is defined as $\Gamma(p) = \{(i, j) \mid (p_i + p_j)/2 = p\}$. The symmetry map $M_{Gray}(p)$ is a sum over all pixel pairs that surround p:

$$M_{Gray}(p) = \sum_{(i,j)\in\Gamma(p)} PWF_{Gray}(i, j) \cdot GWF(i, j) \cdot DWF_\sigma(i, j). \quad (1)$$

The functions PWF_{Gray}, GWF and DWF_σ evaluate how strong a pixel pair (p_i, p_j) contributes to symmetry. Here, the *Phase Weight Function* PWF_{Gray} is a measure for the probability that gradient directions at p_i and p_j belong to a symmetric object:

$$PWF_{Gray}(i, j) \;=\; (1 - \cos(\gamma_i + \gamma_j)) \;\cdot\; (1 - \cos(\gamma_i - \gamma_j))\,. \quad (2)$$

γ_i, γ_j denote the angle between the local gradients at p_i and p_j, respectively, and the line $\overline{p_i p_j}$ connecting p_i and p_j. If α_{ij} denotes the angle between $\overline{p_i p_j}$ and the horizon, then $\gamma_i = \theta_i - \alpha_{ij}$, $\gamma_j = \theta_j - \alpha_{ij}$. PWF_{Gray} takes high values if the gradients at p_i and p_j are directed such that they belong to the contours of an object which is symmetric around the central point $p = (p_i + p_j)/2$. For a detailed discussion on PWF_{Gray} see [12, 6].

The second factor in Eq. 1 is the *Gradient Weight Function* GWF which weights the contribution of pixels (p_i, p_j) higher when both of them are located on edges:

$$GWF(i,j) \;=\; \log(1 + G_I(p_i)) \;\cdot\; \log(1 + G_I(p_j)). \tag{3}$$

The idea is that edges are likely to be borders of an object and thus more relevant for symmetry detection. The logarithm attenuates the influence of very strong edges. In the implementation used here, computational efficiency is improved by evaluating only pixel pairs p_i, p_j where both pixels are located on a sufficiently strong edge, as proposed in [10].

The third factor in Eq. 1 is the *Distance Weight Function* DWF_σ:

$$DWF_\sigma(i,j) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma^2}\right). \tag{4}$$

The parameter σ defines the scale on which symmetries are detected. Since the only effect of DWF is a slight smoothing of the saliency map M_{Gray}, it will be left out in the implementation used here. To maintain the locality of the symmetry measure, the summation of the pixel pairs defined by Γ is now restricted to pairs with $\|p_i - p_j\| \leq 2R$, where the parameter $R > 0$ will be called the "symmetry radius". In other words, the new version of Γ is

$$\Gamma(p) = \{(i,j) \mid (p_i + p_j)/2 = p \;\wedge\; \|p_i - p_j\| \leq 2R\}\,. \tag{5}$$

This modification has no major effect since the purpose of DWF_σ is to make the measure local, i.e. contributions of pixel pairs further away than $\approx 3\sigma$ from the central pixel p become very small. The same can be achieved by a circular mask of radius R. Apart from that, DWF_σ leads to a slight smoothing of M_{Gray}, which is achieved in the current implementation by convolution with a Gaussian kernel *after* computation. After the convolution, IPs are detected as the highest maxima.

2.2 The COL-SYM algorithm

2.2.1 Motivation

The GRAY-SYM detector suffers from two problems:

1. In the presence of low contrast, symmetric shapes are often difficult to detect (Fig. 1).
2. The Phase Weight Function PWF_{Gray} is 2π-periodic in the gradient directions. Therefore gradients from dark to bright and gradients from bright to dark have opposite signs, which makes it impossible to detect, e.g., a symmetric gray object on dark background to its left and bright background to its right.

The color based symmetry detector (COL-SYM) introduced in [6] overcomes both shortcomings by operating equally on all three channels. As COL-SYM is an extension of GRAY-SYM, it is also based on gradients. The problem that arises with color is that the extension of the gray value gradient to color is a tensor, which can not be easily used for symmetry detection. Therefore, gradients are still computed for the isolated color channels, but symmetry is detected by an across-channel evaluation. In the course of this scheme, a novel phase weight function is introduced which also solves the afore mentioned problem of GRAY-SYM for the single channel case.

Figure 1: Comparison of the detectors GRAY-SYM and COL-SYM. Upper left: Saliency map of GRAY-SYM. Upper right: IPs obtained as maxima from this saliency map. Lower left: Saliency map of COL-SYM. Lower right: IPs for COL-SYM. COL-SYM catches far more of the symmetries in low contrast areas by exploitation of color (e.g. colored pearls). In addition, the new phase weight function leads to higher and easier to detect maxima in the saliency map. For example, the eyes are visible in both saliency maps, but to weak to be detected with a reasonably high threshold for GRAY-SYM.

2.2.2 Description of the algorithm

Let the color image be given by color values $I_i(p)$, where $i = 0, 1, 2$ denotes the red, green and blue channel. Further, let $G_i(p)$ and $\theta_i(p)$ denote the gradient magnitude and gradient direction for each channel. The saliency map $M_{Col}(p)$ is defined as follows:

$$M_{Col}(p) = \sum_{(i,j)\in\Gamma(p)} \sum_{(k,l)\in\Lambda(p,i,j)} PWF_{Col}(i,j,k,l) \cdot GWF_{Col}(i,j,k,l), \tag{6}$$

The main difference to $M_{Gray}(p)$ is the additional summation over all significant color edge pairs, i.e. edge pairs red-red, red-green etc.

To be more precise, the set $\Lambda(p,i,j)$ is the set of pairs of color indices $(k,l),\ k,l \in [0,2]$, for which the gradients exceed predefined thresholds ϑ_k, ϑ_l:

$$\Lambda(p,i,j) = \{(k,l) \ \mid \ k,l \in \{0,1,2\} \ \wedge \ G_k(p_i) \geq \vartheta_k \ \wedge \ G_l(p_j) \geq \vartheta_l\}. \tag{7}$$

In the across-channel summation over $\Lambda(p,i,j)$, color channels have to be treated equally. For example, the contribution of a red-green edge must be equal to the contribution of a green-red edge. Therefore, the 2π-periodic phase weight function PWF_{Gray} (Eq. (2)) must be replaced by

$$PWF_{Col}(i,j,k,l) \ = \ \left[\cos^2(\gamma_{ik} + \gamma_{jl})\right] \ \cdot \ \left[\cos^2(\gamma_{ik}) \cdot \cos^2(\gamma_{jl})\right], \tag{8}$$

where γ_{ik} denotes the angle between gradient $G_k(p_i)$ and the line $\overline{p_i p_j}$. Note both factors are π-periodic, so PWF_{Col} is invariant to transformations $\gamma_{ik} \to \gamma_{ik} + \pi$ (i.e. $PWF_{Col}(\gamma_{ik}, \gamma_{jl}) = PWF_{Col}(\gamma_{ik} + \pi, \gamma_{jl}) = PWF_{Col}(\gamma_{ik}, \gamma_{jl} + \pi) = PWF_{Col}(\gamma_{ik} + \pi, \gamma_{jl} + \pi)$.

The gradient weight function GWF_{Col} is analogous to its gray value version:

$$GWF_{Col}(i,j,k,l) \ = \ \log(1 + G_k(p_i)) \ \cdot \ \log(1 + G_l(p_j)). \tag{9}$$

The resulting map M_{Col} is slightly smoothed before IPs are detected as the N highest maxima (Fig. 1). As for GRAY-SYM, COL-SYM has only one main parameter: the symmetry radius R, which selects the scale on which symmetries are evaluated and thus the size of structures or objects that may be represented by an IP. The parameters ϑ_k are of minor importance to the results but decisive for computational efficiency, as they define the edge strength required for processing. Usually, ϑ_k is chosen equal for all channels in the RGB case.

2.2.3 Color spaces

The above description of COL-SYM was for three channels, but it can easily be generalized to n channels. For example, multispectral images can be processed.

While there is a lot of discussion on color spaces for image processing, experience with applications (as outlined in the next section) shows that the choice of a suitable color space is entirely application specific. For the problem shown in Fig. 2, the RGB-space is well suited due to the colors of the objects of interest. But then, in this setting constant illumination is provided. For tasks where illumination is subject to change, more robust or invariant color spaces are to be preferred.

Choosing a color space appropriately of a specific task is feasible only for restricted, fixed domains, e.g., if the set of objects is small and a priori known, and / or the illumination is fixed. But for applications such as image retrieval, where a huge variety of images must be dealt with, there is little reason to prefer one color space over the other since they yield different sets of IPs. In this case, the parallel use of several color spaces may be sensible, as discussed below.

3 Applications

The idea of using IPs in an image processing architecture is to filter out relevant image areas. Therefore the role of IPs is twofold: (*i*) Computational efficiency is achieved by leaving out irrelevant areas, and (*ii*) pattern recognition is facilitated by the pre-selection of a certain type of patterns (here: symmetric patterns). But due to the relatively simple and close-to-signal nature of IP-algorithms, the power of this filtering process is limited. An IP-detector does no more than providing a set of areas that have an increased probability of being relevant by showing a particular pattern type. Most IP-detectors provide also a measure for this probability, i.e. the "goodness" of an IP. In the case of the symmetry detectors, the height of the maximum in the symmetry map is a good indicator. In an IP-based architecture, a suitable threshold for the required goodness must be found as a trade-off between maximizing true positives (IPs that are where they should be) and minimizing false positives (IPs in places without sufficient symmetry).

3.1 Object and component recognition

For object recognition tasks, IPs have to localize the objects of interest. GRAY-SYM or COL-SYM make sense only when the objects are symmetric, but once there are objects with other prominent features, also other types of IP-detectors have to be applied. If necessary, several IP-detectors can be used in parallel as described in the next section.

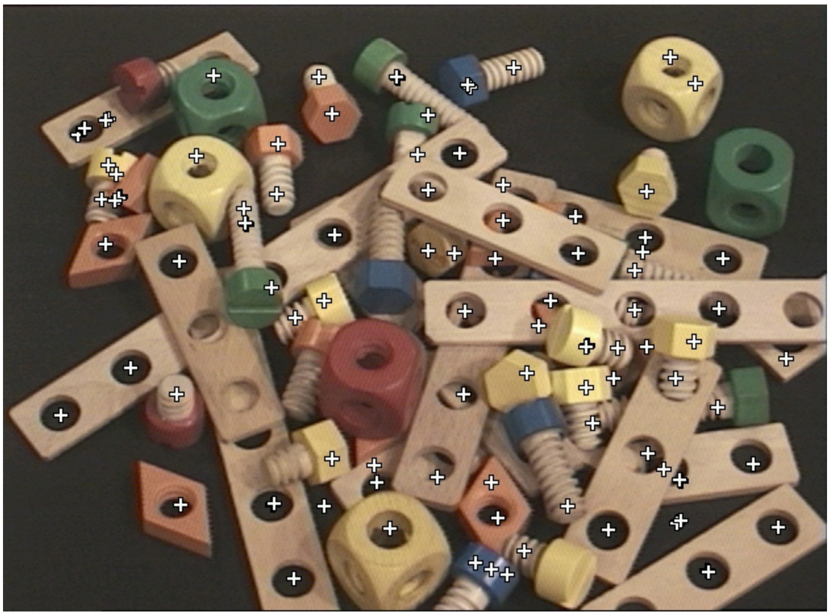

Figure 2: Application of the IP-detector: Symmetric areas are selected and subsequently classified by an object recognition system with categories such as "bolt head" or "hole of a bar". Note that the IP selection is part of the pattern recognition as it restricts possible categories to symmetric objects.

Fig. 2 shows a pile of Baufix-objects. Baufix is a set of wooden toy pieces that can be connected by bolts. In this case, not only the objects are symmetric, but they bear also several symmetric components such as holes or the symmetric heads of bolts. Therefore, a component-based scheme was developed for recognition rather than a classifier for isolated objects. By this means, the objects can be recognized even under strong occlusion (e.g. as parts of aggregates), because it is likely that at least some components of the object remain visible.

Further processing is restricted to square windows centered at the IPs, the rest of the image is now ignored. The windows are classified into categories such as "bolt head", "hole of a bar" or "no known category" by a neural classifier, so the result is a collection of labeled IPs. The next step is to derive an object classification from the classified components and to associate the appropriate image regions. This medium-level processing step is performed by a knowledge based approach [13] which integrates geometrical knowledge about the objects with the information obtained from the low-level.

Summarizing, the role of IPs in object recognition is to provide a set of candidate windows centered at relevant objects or components. These are to be classified and "assembled" by further processing levels, in particular, windows containing unknown or irrelevant structures have to be sorted out. The idea was explained here for a system of toy pieces, another application related to object recognition is in the context of an augmented reality system, where object knowledge is taught online using a mobile interface [2].

3.2 Image retrieval

The use of IPs in content based image retrieval (CBIR) is different from object recognition. This is caused by the fact that the majority of CBIR systems is aimed at *global* retrieval: A set of query images is given by the user, for which the CBIR system returns the most similar images of a database. Similarity is measured by features extracted globally from the query images and the returned images. "Global" does not necessarily imply that the complete image is evaluated — there may be a selection of relevant areas by segmentation or IP-techniques — but the user does not specify which parts of the image are relevant. For this reason, IP-based CBIR systems usually employ the windows selected by the IP-algorithm for a characterization of the complete image, examples are [14, 17]. An overview of CBIR can be found in [16].

In contrast to global retrieval, an entirely *local* CBIR system was proposed in [5]. In this system, for both the query image and the images of the database IPs are computed using several detectors in parallel, including COL-SYM (the older system described in [5] still uses GRAY-SYM, but here the updated version is described). A query is defined by selecting IP-centered windows from the query images, the remainder of the images is completely ignored. Retrieval is performed by comparing the query windows to windows found in the database by the IP-detectors. By this means, the search is restricted to the objects the user wants to find, i.e., if e.g. a yellow flower is selected, then only images with yellow flowers are to be returned, regardless of any background. The principle of work is illustrated in Fig. 3.

Locality of the search and restriction to user-defined windows makes the retrieval task somewhat similar to object recognition. Again, the role of IPs is not only to achieve computational efficiency, but also pattern recognition. For CBIR, IPs define the image components that can be searched for. This is a restriction compared to global approaches, which allow to search for arbitrary patterns, at least in principle. But the restriction to IPs forces the user to specify only patterns that can actually be represented and searched for, which leads to a higher success rate. In case the detected IPs prove to be an insufficient representation, IP detectors covering a greater variety of (color–) features have to be integrated.

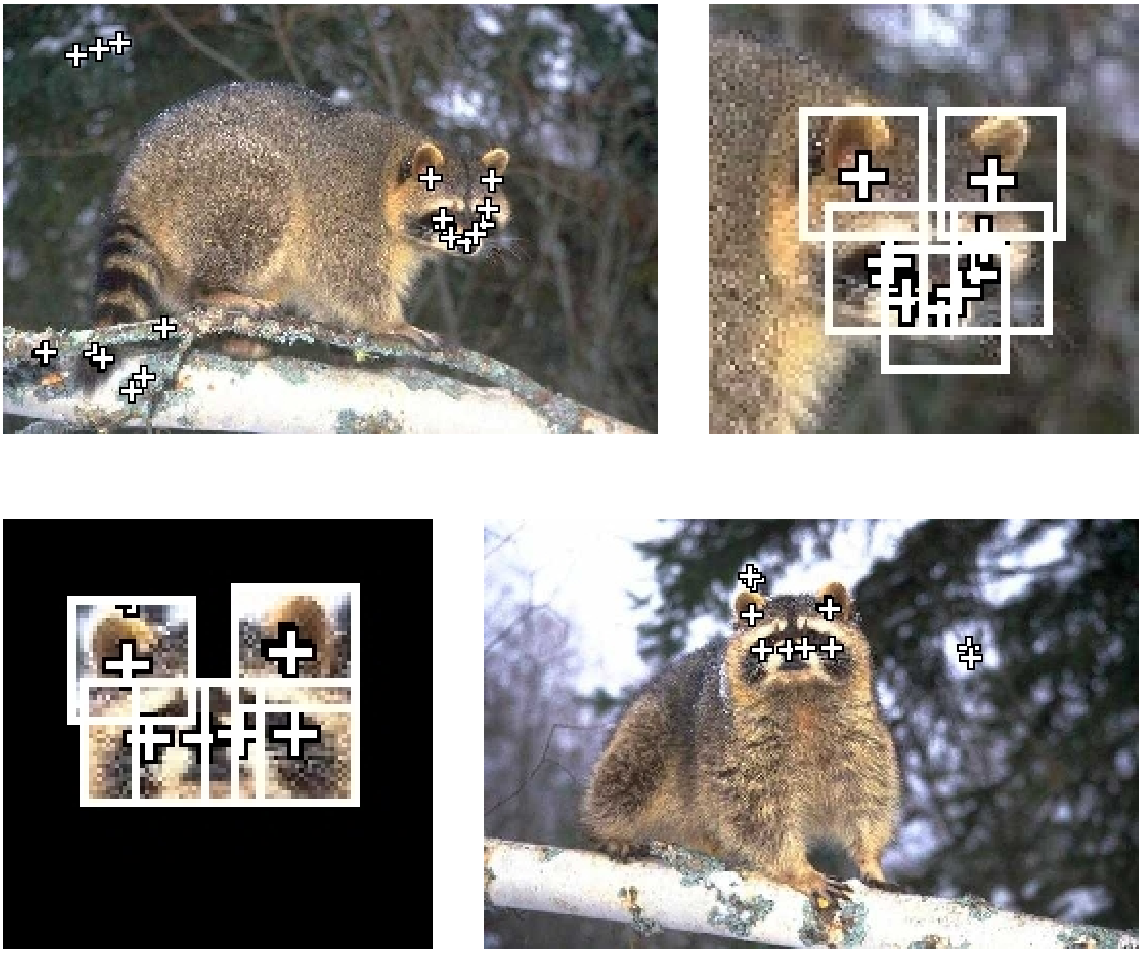

Figure 3: IPs for a retrieval task. First row: In the query image, several IPs have been found, some on the relevant object, some on background (left). The user selects the relevant IPs and the appropriate window size (right). Second row: The search is for windows similar to the specified ones such windows are shown in the left image. The right image shows the complete version of the left.

4 Conclusion

An IP-detector based on symmetry computation was presented which differs from earlier work by the exploitation of color as a form feature. To date there are surprisingly few works on the use of spatio-chromaticity in image processing, which is anachronistic considering the availability of even low cost color cameras. The work presented here is restricted to close-to-signal features, since IPs are a low level concept. But there is no apparent reason why spatio-chromaticity should not be used for higher processing stages also. Therefore, the current work on IPs is merely a first step towards architectures that rely equally on intensity and color on all processing levels.

References

[1] G. Backer, B. Mertsching, and M. Bollmann. Data- and Model-Driven Gaze Control for an Active-Vision System. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(12):1415–1429, 2001.

[2] H. Bekel, G. Heidemann, and H. Ritter. Interactive Image Data Labeling Using Self-Organizing Maps in an Augmented Reality Scenario. *Neural Networks*, 18(5/6):566–574, 2005.

[3] H. D. Cheng, X. H. Jiang, Y. Sun, and J. Wang. Color image segmentation: advances and prospects. *Pattern Recognition*, 34(12):2259–2281, 2001.

[4] C. Harris and M. Stephens. A Combined Corner and Edge Detector. In *Proc. 4th Alvey Vision Conf.*, pages 147–151, 1988.

[5] G. Heidemann. Combining spatial and colour information for content based image retrieval. *Computer Vision and Image Understanding, Special Issue on Colour for Image Indexing and Retrieval*, 94(1-3):234–270, 2004.

[6] G. Heidemann. Focus-of-Attention from Local Color Symmetries. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26(7):817–830, 2004.

[7] G. Heidemann. The Long-Range Saliency of Edge- and Corner-Based Salient Points. *IEEE Trans. on Image Processing*, 14(11):1701–1706, 2005.

[8] G. Heidemann. The Principal Components of Natural Images Revisited. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28(5):822–826, 2006.

[9] D. Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* Freeman Publishers, San Francisco, 1982.

[10] T. W. Nattkemper. Untersuchung und Erweiterung eines Ansatzes zur modellfreien Aufmerksamkeitssteuerung durch lokale Symmetrien in einem Computer Vision System. Master's thesis, Bielefeld Univ., Technische Fakultät, 1997.

[11] N. Pal and S. Pal. A Review on Image Segmentation Techniques. *Pattern Recognition*, 26(9):1277–1294, 1993.

[12] D. Reisfeld, H. Wolfson, and Y. Yeshurun. Context-Free Attentional Operators: The Generalized Symmetry Transform. *Int'l J. of Computer Vision*, 14:119–130, 1995.

[13] G. Sagerer, C. Bauckhage, E. Braun, G. Heidemann, F. Kummert, H. Ritter D., and Schlüter. Integrating recognition paradigms in a multiple-path architecture. In S. Singh, N. Murshed, and W. Kropatsch, editors, *Proc. Int'l Conf. on Advances in Pattern Recognition (ICAPR)*, pages 202–211, Rio de Janeiro, 2001. Springer.

[14] C. Schmid and R. Mohr. Local Grayvalue Invariants for Image Retrieval. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(5):530–535, 1997.

[15] C. Schmid, R. Mohr, and C. Bauckhage. Evaluation of Interest Point Detectors. *Int'l J. of Computer Vision*, 37(2):151–172, 2000.

[16] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-Based Image Retrieval at the End of the Early Years. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, 2000.

[17] Q. Tian, N. Sebe, M. S. Lew, E. Loupias, and T. S. Huang. Image Retrieval Using Wavelet-Based Salient Points. *J. of Electronic Imaging*, 10(4):835–849, 2001.