Focus-of-Attention from Local Color Symmetries

Gunther Heidemann

Abstract—In this paper, a continuous valued measure for local color symmetry is introduced. The new algorithm is an extension of the successful gray value-based symmetry map proposed by Reisfeld et al. The use of color facilitates the detection of *focus points* (FPs) on objects that are difficult to detect using gray-value contrast only. The detection of FPs is aimed at guiding the attention of an object recognition system; therefore, FPs have to fulfill three major requirements: stability, distinctiveness, and usability. The proposed algorithm is evaluated for these criteria and compared with the gray value-based symmetry measure and two other methods from the literature. Stability is tested against noise, object rotation, and variations of lighting. As a measure for the distinctiveness of FPs, the principal components of FP-centered windows are compared with those of windows at randomly chosen points on a large database of natural images. Finally, usability is evaluated in the context of an object recognition task.

Index Terms—Focus-of-attention, color vision, symmetry, saliency maps, object recognition.

1 Introduction

Selection of salient image points and regions as candidates for further processing is a key problem in many fields of image processing like object recognition [20], [51], content-based image retrieval [56], [8], [64], [62], active vision [1], [4], [10], [13], [3], image compression [50], or medical image registration [40]. There are mainly two different approaches: *Goal-driven* techniques search exclusively for regions which have a high probability to belong to a certain object, an example is eigenspace methods which allow a highly selective localization of faces [43]. While goal-driven techniques are designed for a special task, *context-free* methods are not specialized on the search for a certain object. Instead, they judge the relevance of image points from the "saliency" of signal features alone. As a consequence, context-free methods have a wider range of applicability and can be used to start the bottom up-top down processing cycle, however, at the cost of less selectivity.

Most algorithms for context-free salient point detection are designed to detect corners or edges in gray-value images [44], [21], [56], [60], [66], [67]. Another type of methods is based on previously extracted contours to find, e.g., points of high curvature or intersection points [2], [42], [28], [58]. However, corners and edges often indicate the border of an object, not the center, which is needed for many object recognition algorithms.

1.1 Salient Points and Focus Points

In this paper, a distinction is drawn between *salient points* (or *interest points* in terms of [56]) and *focus points*. Salient points (SPs) are understood as pixels that are "interesting" because of the signal features within a *small* neighborhood. Therefore, all of the above-mentioned corner and edge-based algorithms are aimed at detecting SPs. However, as SPs indicate only the "interestingness" of very small image patches, these patches often don't correspond to interesting larger entities like physical objects.

In contrast, a focus point (FP) will be understood as the center of a larger region which is 1) interesting *as a whole* and which is 2) *usable for recognition*. So, for FPs, the emphasis is on their *function* within a vision system: FPs can serve as the center of a window W from which features can be extracted for object recognition. W can be thought of as an object centered reference frame suitable for recognition in terms of Palmer [49], however, without being the result of the search for a special object. The size of W depends on the particular application and is not necessarily related to the scale on which saliency was detected. For example, algorithms for corner detection (see Section 1.4) might direct attention to a big object requiring a large window by means of a small detail (corner). In contrast, a detector for large-scale symmetries will detect large symmetric objects, though the application may require smaller windows to recognize object details.

1.2 The Role of Focus Points in View-Based Object Recognition

Object recognition is a difficult problem since the pixel appearance of an object is subject to substantial changes due to its various "visual degrees of freedom" (DOF): Translation (3 DOF), rotation (3 DOF), lighting conditions (number and location of sources, brightness, spectra), background, partial occlusion, and, if so, nonrigidness of the object. Considering the image patch covered by the object as a point in a high-dimensional space P spanned by the gray or color values of its pixels, appearance variability causes an object to cover a complicated "appearance manifold" $A \subset P$. The *view-based* approach to object recognition memorizes a representation of A based on samples. *Recognition* means comparing the pixel appearance of an unknown object with the memorized representations of the A_i ($i = 1 \ldots N$ for N objects).

FPs are a means to simplify the appearance manifold A and, thus, facilitate its representation: In an object recognition architecture, FP-detection rules out all but a few points as centers for feature extraction. In other words, FP-detection solves part of the well-known "what-where-problem," i.e., the problem of identifying an object's location and identity

• *The author is with the Faculty of Technology, Neuroinformatics Group, University of Bielefeld, PO Box 100131, D-33501 Bielefeld, Germany. E-mail: gheidema@techfak.uni-bielefeld.de.*

Manuscript received 25 Oct. 2002; revised 14 May 2003; accepted 26 June 2003.
Recommended for acceptance by W. Freeman.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 117666.

simultaneously: The where-problem is reduced by the use of FPs because evaluation is restricted to a few image patches and the what-problem is partly solved because only certain image patches are selected by saliency criteria.

1.3 Criteria for FP-Detectors

To fulfill the role in object recognition outlined above, FPs must meet certain requirements:

1. **Stability:** FP-positions on objects must be stable against object rotations over a reasonable angular range. If an FP disappears under object rotation, there has to be a stable new one on the object at another location. Moreover, FPs must be robust against noise and changes in lighting.
2. **Distinctiveness:** An FP must represent the center of a window which is "salient" or "distinctive" in the sense that the probability to find such a local pattern is small compared to other, more frequent patterns of natural images.
3. **Usability:** FPs serve as centers of windows for feature extraction, therefore, locations must be such that the windows overlap largely with the objects of the domain or with other suitable "visual entities," like certain parts of objects. However, it is not required that *all* FPs are well-positioned. The task to distinguish object centered FPs from others—e.g., at the borders of objects—remains for the subsequent classification system.

These criteria reflect two of the main properties which make a point interesting according to Haralick and Shapiro [20]: invariance and distinctiveness. In contrast to [20], the criteria used here do not include that an FP must be *distinguishable* from its immediate neighbors—this requirement inevitably favors edges and corners. Therefore, only the distinctiveness of the window around the FP compared to other (random) windows is required here. Moreover, global uniqueness is not required because this would forbid the simultaneous detection of several objects of the same type.

1.4 Algorithms for Focus-of-Attention

Most works on salient point (SP) detection concentrate on edges or corners. The computation often relies on discrete approximations of the autocorrelation function of the signal [21], [16], [63], as proposed originally by Moravec [44]. As an example, Figs. 5 and 7 show SPs found by the detector of Harris and Stephens [21] (see the Appendix). Many methods rely on first and/or second derivatives of the signal [6], [33], [15] or evaluate previously extracted contours [2], [42], [28], [58]. Lee et al. use wavelets for corner detection [36], [11]. An alternative to filter-based methods is presented by Laganiere, who detects L-shaped corners using morphological operators [35]. The SUSAN-detector introduced by Smith and Brady finds corners and edges using form features of the region of pixels with gray values similar to the central pixel within a circular mask [60].

One of the most simple approaches to compute saliency from image features other than corners and edges is entropy calculation of the gray-value histogram of image windows ([31], for an application see [19]). Since this procedure is related to local gray-value variance computation, textured regions are highly favored. To detect larger, "blob-like" visual entities an approach was introduced by Lindeberg [37], which is based on scale space theory [38].

Application of saliency detectors as components of real image processing architectures can be found in the area of active vision. Rao et al. [52] introduced a model for visual search using steerable filters [17]. Approaches for the integration of *several* elementary saliency features for gaze control were proposed by Itti et al. [30] and Backer et al. [3]. Both architectures rely on findings about the primate visual system and are highly domain-independent. Shokoufandeh et al. use saliency maps in an entirely different way: They define graphs between detected salient object regions and calculate the difference to memorized graphs for recognition [59].

One of the most promising approaches to context-free saliency detection is the evaluation of local symmetries proposed by Reisfeld et al. [53], which will be outlined in the next section. The use of symmetry is motivated by psychophysical findings [32], [9], [39] which indicate that "symmetry catches the eye" as stated by Locher and Nodine [39]. Further evidence for the approach of Reisfeld et al. was found by Privitera and Stark [50], who found a high correlation of the predicted SPs with human eye fixations for general images. Another continuous symmetry measure proposed by Zabrodsky et al. [65] yields similar advantages, but, as it depends strongly on a preceding contour extraction, it was not considered here.

1.5 Contents of This Paper

Based on the algorithm proposed by Reisfeld et al. [53], a novel method for the continuous, context-free judgment of local color symmetries will be developed in Section 3. To compare both methods, it will be checked in Section 4 whether the three requirements to FPs set up in Section 1.3 are fulfilled. Stability against rotations is tested using the Columbia Object Image Library (COIL), robustness against noise is tested on images from a low-cost camera, and the influence of lighting is evaluated for four different illuminations. Saliency is judged by comparing the principal components of FP-centered windows of natural images with those of randomly sampled windows on a database of over 89,000 images. For comparison, the well-known SP-detector of Harris and Stephens [21] will be tested as a representative of corner and edge-based methods and a detector using Daubechies4 wavelets [14] proposed by Tian et al. [62].

Usability of FPs can be judged only in the framework of an actual recognition system. Therefore, in Section 5, the recognition of assembled wooden toy pieces is presented as a sample application.

2 The Original Approach of Reisfeld et al.

2.1 Construction of the Symmetry Map

The algorithm proposed by Reisfeld et al. [53] calculates a continuous, local symmetry judgment based on gray-value edges. First, the original version of the isotropic symmetry measure M_{Org} will be outlined. Notation is slightly different from [53] for later convenience. 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 $(I_x(p), I_y(p))$, from which the gradient magnitude $G_I(p)$ and direction $\theta_I(p)$ can be calculated (horizon is $\theta = 0$):

$$I_x(p) = \frac{\partial I(p)}{\partial x}, \quad I_y(p) = \frac{\partial I(p)}{\partial y},$$
$$G_I(p) = \sqrt{I_x(p)^2 + I_y(p)^2}, \theta_I(p) = \arctan\left(\frac{I_y(p)}{I_x(p)}\right). \tag{1}$$

For each pixel p, a set $\Gamma(p)$ of index pairs (i,j) of surrounding pixel pairs (p_i, p_j) is defined by

$$\Gamma(p) = \left\{(i,j) \;\middle|\; \frac{p_i + p_j}{2} = p\right\}, \tag{2}$$

(see Fig. 1a). The isotropic symmetry map $M_{Org}(p)$ is a sum over all pixel pairs surrounding p:

$$M_{Org}(p) = \sum_{(i,j)\in\Gamma(p)} PWF_{Gray}(i,j) \cdot GWF(i,j) \cdot DWF_\sigma(i,j). \tag{3}$$

The contributions of the functions PWF_{Gray}, GWF, and DWF_σ will be discussed in detail.

2.1.1 Original Phase Weight Function PWF_{Gray}

The most important contribution is the gray value-based *Phase Weight Function* PWF_{Gray}, which consists of two factors:

$$PWF_{Gray}(i,j) = \underbrace{\left[1 - \cos(\gamma_i + \gamma_j)\right]}_{PWF^+_{Gray}(i,j)} \cdot \underbrace{\left[1 - \cos(\gamma_i - \gamma_j)\right]}_{PWF^-_{Gray}(i,j)}. \tag{4}$$

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, see Fig. 1b. Let α_{ij} denote the angle between $\overline{p_i p_j}$ and the horizon, so

$$\gamma_i = \theta_i - \alpha_{ij}, \quad \gamma_j = \theta_j - \alpha_{ij}. \tag{5}$$

PWF_{Gray} takes a high value if the gradients at p_i and p_j are directed such that they might be part of the contours of an object which is symmetric around the central point $p = (p_i + p_j)/2$. This is achieved by two factors:

$\mathbf{PWF^-_{Gray}}$: $PWF^-_{Gray}(i,j)$ is maximal for opposite gradient directions ($\gamma_i - \gamma_j = \pm 180°$) which indicate a bright/dark object at p on dark/bright background, respectively (Fig. 1d). $PWF^-_{Gray}(i,j)$ is minimal for $\gamma_i = \gamma_j$, which indicates that p_i and p_j are on the same side of an object (Fig. 1e). The transition between these extremes is continuous, thus allowing nonperfect symmetries. However, as shown in Figs. 1d, 1e, 1f, and 1g, this criterion indicates symmetry only if the condition imposed by PWF^+_{Gray} is fulfilled simultaneously.

$\mathbf{PWF^+_{Gray}}$: $PWF^+_{Gray}(i,j)$ is a measure for the symmetry of the two gradients with respect to the perpendicular bisector of $\overline{p_i p_j}$ (Figs. 1d, 1e, 1f, and 1g). Again, a bright object on dark background is treated like a dark object on a bright background.

2.1.2 Gradient Weight Function GWF

The *Gradient Weight Function* GWF is aimed to weight the contribution of pixels (p_i, p_j) higher if they are both on edges because edges might relate to object borders:

$$GWF(i,j) \;=\; \log(1 + G_I(p_i)) \;\cdot\; \log(1 + G_I(p_j)), \tag{6}$$

with G_I defined in (1). The logarithm attenuates the influence of very strong edges.

2.1.3 Distance Weight Function DWF_σ

The *Distance Weight Function* DWF_σ makes the symmetry $M_{Org}(p)$ a local measure:

$$DWF_\sigma(i,j) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma^2}\right). \tag{7}$$

The choice of σ defines the scale on which symmetries are detected.

2.2 Implementation of the Symmetry Map

In principle, for each pixel p the saliency $M_{Org}(p)$ has to be summed up over the entire set Γ. However, the computational effort can be reduced greatly by minor modifications without substantial influence on the resulting map:

1. **Restriction to edge evaluation:** Since edges give the main symmetry contribution due to GWF, computations can be restricted to edges only, as proposed by Nattkemper [46], [24]. x and y-gradients are calculated by Sobel filtering, then, from the two maps, the gradient magnitude G_I is calculated and, from G_I, a thresholded binary version G_I^B. All computationally more expensive steps are carried out only for pixels where G_I^B is 1. So, instead of Γ, a restricted set $\Gamma' = \Gamma \setminus \{(i,j) \mid G_I^B(p_i) = 0 \vee G_I^B(p_j) = 0\}$ is used.
2. **Restriction to circular area:** As DWF_σ practically excludes contributions of pixels p' with distance $\|p - p'\| > 3\sigma$, Γ' can be further restricted to $\Gamma^* = \Gamma' \setminus \{(i,j) \mid \|p_i - p_j\| > 2R\}$, where R will be called the "symmetry radius."
3. **Dispose of** DWF_σ: The Gaussian weighting within the surroundings of p has a slightly smoothing effect on the resulting symmetry map. However, the effect of DWF_σ on the resulting FP-positions can be neglected if the map is postprocessed by convolution with a Gaussian. Therefore, DWF_σ is left out, so now the symmetry radius R alone defines the "scale of interest" instead of σ.

Hence, (3) can be replaced by the more efficient version

$$\boxed{M_{Gray}(p) = \sum_{(i,j)\in\Gamma^*(p)} PWF_{Gray}(i,j) \cdot GWF(i,j).} \tag{8}$$

To obtain FPs, M_{Gray} is smoothed by convolution with a Gaussian kernel. The FPs are then the N_{FP} highest maxima. This entire procedure for FP-detection will be referred to as GraySym.

To check that the influence of the modifications is negligible, FPs were generated on 30 randomly chosen natural images of the Art Explosion® Photo Gallery [48] using both the original (3) and the modified version (8). As long as $R \geq 3\sigma$ holds, there are no substantial differences: It was checked that the FPs generated from the 10 best maxima of the symmetry map did not change their pixel positions. Hence, contributions of pixels farther than 3σ from the center are not essential for FP-detection.

3 Saliency Maps from Color Symmetries

In the following, a new symmetry measure based on color edges will be described. The idea is to exploit not only gray-value contrast but also color contrast. As an example, Fig. 2

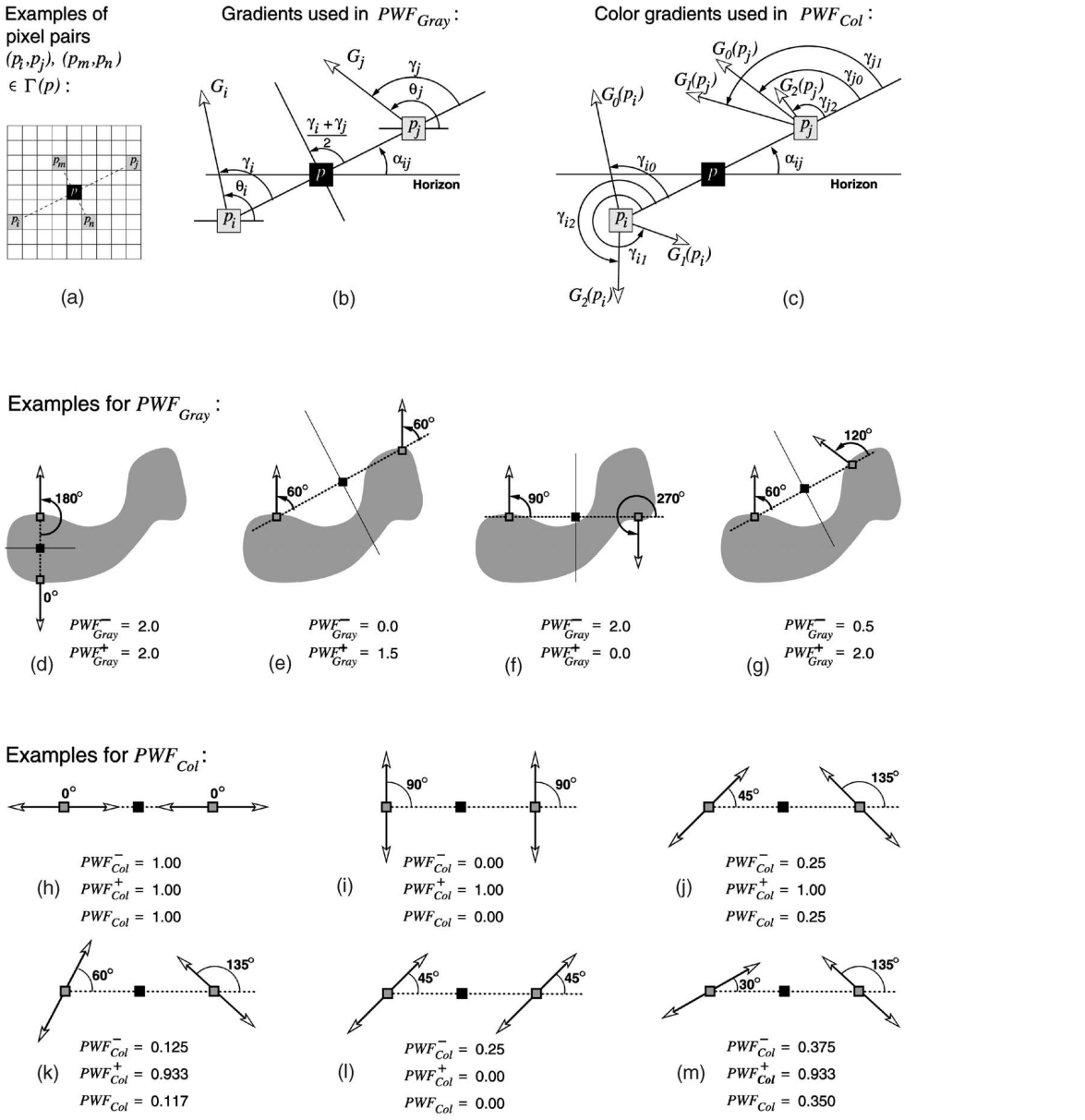

Fig. 1. Upper row: Nomenclature of gradients and angles for PWF_{Gray} and PWF_{Col}. Below some examples for different gradient angles. As PWF_{Col} is invariant to changes of gradient direction by 180 degrees (in contrast to PWF_{Gray}), both possible directions—which lead to the same result—are indicated by the double arrows.

shows colored squares on a background of equal gray-value luminance: M_{Gray} does not reflect the symmetries as the edges are invisible in the gray value-based edge map, whereas the color symmetry map M_{Col} proposed in Section 3.2 indicates all symmetries visible within the chosen radius R. As a motivation for the more difficult calculation of M_{Col}, in Section 3.1 first a simple extension of the gray value-based measure (8) will be discussed.

3.1 Motivation: A Simple extension of Gray-Value Symmetry

As a "naive" solution to detect objects which differ from the background in color but not in gray values, Fig. 2 shows a "colored" symmetry map $M^{3\to3}_{Col}$. This map is the straightforward extension of the gray-value map: (8) is applied to the R, G, and B channels separately. So, the three input channels are mapped to three output channels ("$3 \to 3$"), which can be interpreted as "colored" symmetry map. In Fig. 2, the resulting colored symmetry map $M^{3\to3}_{Col}$ shows, e.g., a "yellow symmetry" for the green-on-red square, for the red-on-green or the cyan-on-magenta square, as all of these have edges both in the red channel and in the green channel. To generate FPs, $M^{3\to3}_{Col}$ could be evaluated for maxima in each channel separately.

However, $M^{3\to3}_{Col}$ has an obvious shortcoming: In Fig. 2, a "green symmetry" between the green-on-blue square and the border between the blue and the cyan background is detected (which is correct), but no symmetry between the blue-on-green square and the border between the green and yellow background (as indicated). The reason for this cognitively unsatisfying result is that no blue edge occurs on the green-yellow border. This effect becomes even more clear in Figs. 3a,

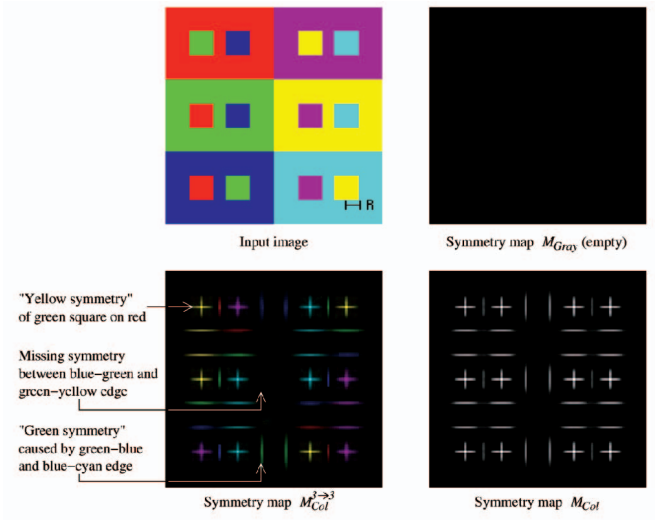

Fig. 2. As the gray values of the squares do not differ from the background, the symmetry is not visible in the symmetry map M_{Gray} after Reisfeld et al. [53]. In contrast, the color edge-based symmetry maps $M_{Col}^{3\to3}$ and M_{Col} indicate the symmetries of the squares and also between the squares and the borders of the background rectangles. Symmetry radius is $R = 12$, edge length of squares $= 20$.

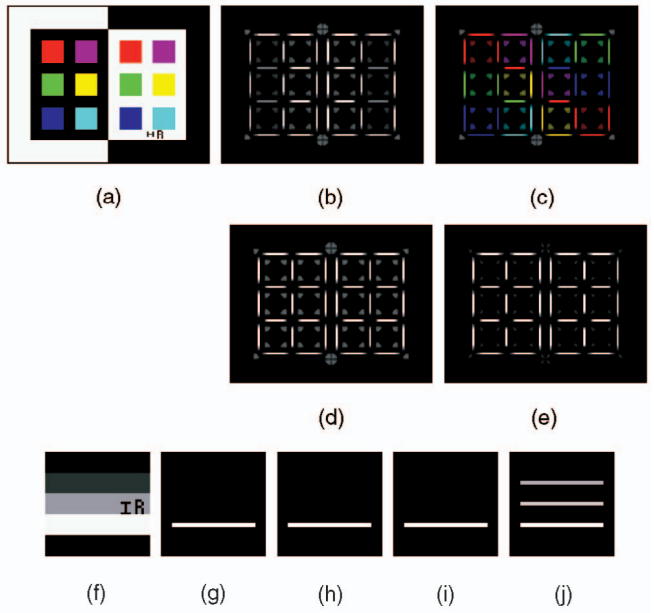

Fig. 3. **Colored squares image** (a), (b), (c), (d), and (e): Symmetry radius $R = 5$, edge length of squares $= 20$, distance between squares $= 10$, so note it is the aim to detect the symmetry of the black or white bars *between* the squares, *not* the symmetry of the squares themselves. M_{Gray} (b) detects all symmetries but differently in strength, $M_{Col}^{3\to3}$ (c) misses some symmetries. M'_{Col}(d) and M_{Col} (e) detect all symmetries. (a) Input image. (b) M_{Gray}. (c) $M_{Col}^{3\to3}$. (d) M'_{Col}. (e) M_{Col}. **Gray bars image** (f), (g), (h), (i), and (j): Symmetry radius $R = 5$, width of the bars $= 10$. Only M_{Col} (j) detects the symmetry of the gray bars because of its π-periodic phase weight function PWF_{Col}, while the PWF_{Gray}-based maps M_{Gray} (g), $M_{Col}^{3\to3}$ (h), and M'_{Col} (i) fail. Image (f) resembles part of a SUSAN test image ([60, p. 55]). (f) Input image. (g) M_{Gray}. (h) $M_{Col}^{3\to3}$. (i) M'_{Col}. (j) M_{Col}.

3b, 3c, 3d, and 3e, where the radius R is chosen such that the symmetries of the black or white bars between the squares are to be detected (not the symmetry of the squares themselves). Here, $M_{Col}^{3\to3}$ detects part of the symmetries but not all: There is, e.g., no symmetry indicated for the black bar between the red and the green square, whereas the symmetry *is* detected for the black bar between the magenta and yellow squares. One might object that the black bars are not "foreground," so their symmetry would not be relevant. But, this fact 1) cannot be recognized on the low level and 2) results should at least be the same for a black bar between red and green and a black bar between magenta and yellow. The symmetry map M_{Col} proposed in the next section will fulfill these requirements.

M_{Gray} *does* detect all symmetries, but differently in strength: The symmetry of the black bar between red and green is weaker than between magenta and yellow.

3.2 The Color Symmetry Map M_{Col}

To avoid the inconsistencies of $M_{Col}^{3\to3}$, a color symmetry map M_{Col} will be constructed that evaluates not only pairs of color edges *of the same channel* (i.e., red-red, green-green, blue-blue), but also pairs of edges of different color channels (e.g., red-green). By this means, the symmetry of an object, which is confined by edges resulting from varying color transitions, can be detected. M_{Col} maps the three color input channels to only one symmetry value ("$3 \to 1$").

For better understanding, first a preliminary version M'_{Col} will be described: Let the color image be given by color values $I_i(p)$, where $i = 0, 1, 2$ denotes the red, green, and blue channel. Similar to (1), the gradients are defined as

$$\begin{aligned} I_{i,x}(p) &= \frac{\partial I_i(p)}{\partial x}, & I_{i,y}(p) &= \frac{\partial I_i(p)}{\partial y}, \\ G_i(p) &= \sqrt{I_{i,x}(p)^2 + I_{i,y}(p)^2}, \\ \theta_i(p) &= \arctan\big(I_{i,y}(p) \,/\, I_{i,x}(p)\big). \end{aligned} \tag{9}$$

Fig. 1c illustrates the notation. The symmetry values $M'_{Col}(p)$ are calculated like $M_{Gray}(p)$ in (8) but with an additional summation over all significant color edge pairs:

$$\begin{aligned} M'_{Col}(p) = \sum_{(i,j)\in\Gamma^*(p)} \; \sum_{(k,l)\in\Lambda(p,i,j)} & PWF'_{Col}(i,j,k,l) \\ & \cdot GWF_{Col}(i,j,k,l), \end{aligned} \tag{10}$$

where Γ^* and Λ are defined as follows: As in (8), $\Gamma^*(p)$ denotes again the set of index pairs (i, j) of all opposing pixels around p within radius R:

$$\Gamma^*(p) = \Big\{(i,j) \;\Big|\; \frac{p_i + p_j}{2} = p \;\wedge\; \|p_i - p_j\| \le 2R\Big\}. \tag{11}$$

For each pixel pair (p_i, p_j), given by $\Gamma^*(p)$, $\Lambda(p, i, j)$ denotes 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 (note the edge thresholds might be different for each color channel):

$$\begin{aligned} \Lambda(p,i,j) = \Big\{(k,l) \;\Big|\; & k,l \in \{0,1,2\} \\ & \wedge\; G_k(p_i) \ge \vartheta_k \;\wedge\; G_l(p_j) \ge \vartheta_l\Big\}. \end{aligned} \tag{12}$$

For the first version, $M'_{Col}(p)$, the original phase weight function (4) is extended to

$$PWF'_{Col}(i,j,k,l) = \big[1 - \cos(\gamma_{ik} + \gamma_{jl})\big] \cdot \big[1 - \cos(\gamma_{ik} - \gamma_{jl})\big], \tag{13}$$

where γ_{ik} denotes the angle between gradient G_{ik} of color channel k at pixel p_i and the line $\overline{p_ip_j}$ (similar to (4) and(5)). See Fig. 1c for the notation.

By analogy to (6), the gradient weight function becomes

$$GWF_{Col}(i,j,k,l) = \log(1+G_k(p_i)) \cdot \log(1+G_l(p_j)). \quad (14)$$

Fig. 3d shows that M'_{Col} overcomes the inconsistencies of $M^{3\to3}_{Col}$: Symmetries are equally detected for *all* black or white bars between the squares. However, it turns out that the summation over all color edge pairs does not solve all problems. Figs. 3f, 3g, 3h, and 3i show another shortcoming of M_{Gray}, $M^{3\to3}_{Col}$, and also M'_{Col}: Only the symmetry of the white bar can be detected, not of the two gray bars. This is due to the fact that all three symmetry measures use the same 2π-periodic phase weight function $1-\cos(x)$. This choice is the reason that symmetric objects on inhomogeneous background cannot be detected if the background is brighter on one side than the object and darker on the opposite side—which is the case for the gray bar in the middle of Fig. 3f. However, inhomogeneous background is to be expected in natural environments.

Hence, the requirement of Reisfeld et al. [53] that only gradients pointing toward each other or away from each other contribute to the symmetry of the central point will be abandoned by introduction of a new phase weight function

$$PWF_{Col}(i,j,k,l) = \underbrace{\left[\cos^2(\gamma_{ik}+\gamma_{jl})\right]}_{PWF^+_{Col}(i,j,k,l)} \cdot \underbrace{\left[\cos^2(\gamma_{ik})\cdot\cos^2(\gamma_{jl})\right]}_{PWF^-_{Col}(i,j,k,l)}, \quad (15)$$

where now the π-periodic function $\cos^2(x) = \frac{1}{2}(1+\cos(2x))$ is used instead of $1-\cos(x)$. This makes PWF_{Col} invariant to transformations $\gamma_{ik} \to \gamma_{ik}+\pi$, which means PWF_{Col} has the same value for gradients rotated by π:

$$\begin{aligned} PWF_{Col}(\gamma_{ik},\gamma_{jl}) &= PWF_{Col}(\gamma_{ik}+\pi,\gamma_{jl}) && (16)\\ &= PWF_{Col}(\gamma_{ik},\gamma_{jl}+\pi) \\ &= PWF_{Col}(\gamma_{ik}+\pi,\gamma_{jl}+\pi). && (17)\end{aligned}$$

Similarly to the gray-value measure (4), PWF_{Col} consists of two factors which will be discussed separately. Figs. 1h, 1i, 1j, 1k, 1l, and 1m show some examples for the judgment of different gradient directions, the $180°$-invariance is indicated by the double arrows.

- $\mathbf{PWF^-_{Col}}$: $PWF^-_{Col}(i,j)$ is a continuous judgment on how parallel both gradients are to $\overline{p_ip_j}$. It is maximal ($PWF^-_{Col}=1$) if both are parallel. It is minimal ($PWF^-_{Col}=0$) if at least one is perpendicular to $\overline{p_ip_j}$. This factor expresses that only gradients pointing towards or away from the central pixel p can contribute to the symmetry value of p. If gradients perpendicular to $\overline{p_ip_j}$ belong to a symmetric object, they may contribute only to the symmetry values of other pixels but not of p.
- $\mathbf{PWF^+_{Col}}$: $PWF^+_{Col}(i,j)$ judges how well gradients (or the opposites) at p_i and p_j conform to a mirror symmetry through the central pixel p. For example, in Figs. 1h, 1i, and 1j, perfect mirror symmetry to a vertical axis through the central pixel is realized, so $PWF^+_{Col}=1$. Nonperfect symmetry exists in Figs. 1k and 1m, where $PWF^+_{Col}=0.933$. In Fig. 1l, there is no mirror symmetry, so $PWF^+_{Col}=0$—though the gradients are parallel.

Using the π-periodic PWF_{Col}, the final version of the color symmetry map is

$$\boxed{M_{Col}(p) = \sum_{(i,j)\in\Gamma^*(p)} \sum_{(k,l)\in\Lambda(p,i,j)} PWF_{Col}(i,j,k,l)\cdot GWF_{Col}(i,j,k,l)} \quad (18)$$

with Γ^*, Λ, GWF_{Col}, and PWF_{Col} given by (11), (12), (14), and (15), respectively. Again, FPs are detected as the N_{FP} highest maxima from a smoothed version of M_{Col}. This entire procedure of FP-detection will be referred to as **ColSym**.

3.3 Color Representation

ColSym was described for the RGB color space. Since ColSym exploits color based on gradient detection, the use of other color spaces is simple as long as gradient calculation is possible. If necessary, the edge thresholds ϑ_i (12) have to be adjusted. Gradient calculation is straightforward for most color spaces, an exception is, e.g., the HSV representation, where the hue-channel and the singularity require special treatment.

The choice of the color space depends on the application. Since different color representations provide different similarity measures, edges of a certain object may be strong in one color space but weak in another, thus affecting the visibility of its possible symmetries. Therefore, finding the best color space for symmetry detection is a similar task as in the field of color segmentation, where, e.g., objects have to be separated from background by color difference.

4 Evaluation

4.1 Qualitative Discussion of ColSym

Calculation of symmetries profits most from the use of color when a symmetric object differs from its surroundings not in gray values but in color only. Examples are shown in Fig. 4: The eyes of the cat are obviously symmetric, but the contrast is low in the gray-value version (upper row). As a consequence, GraySym detects only the symmetry of the pupils (small white dots in M_{Gray}) and some minor symmetries of the surroundings. In contrast, ColSym detects the symmetry of the green eyes *as a whole*. This leads to a large peak in the symmetry map that can be detected much better.

In Fig. 4, bottom row, a picture of poor quality is shown with low contrast in the gray-value version. Though the color version is not sharper, color contrast leads to the detection of the row of small windows above the arc. Further examples for the superiority of ColSym in the presence of color contrast are show in Fig. 5: In the "Apples" image, most of the 30 highest maxima of M_{Col} are on the highly symmetric apples in the foreground, which are hardly "visible" for GraySym. Instead, FPs are found by GraySym only in the strongly contrasted light-shadow regions under the tree—which appear only as minor maxima in M_{Col} because they are not very symmetric. The leaves are not detected since their symmetry is poor (in the projection to the image plane) and contrast is low both in color and gray values. So, the leaves could be detected only for large values of N_{FP}, as becomes clear from M_{Col}.

For the Lena image, differences between ColSym and GraySym are small due to the good contrast in the gray-value image. The "Flowers" image shows results similar to the "Apples" image. The gray value-based algorithms "Harris" and "Daubechies4" outlined in the next section yield points on corners and edges which are only in few cases on the flowers.

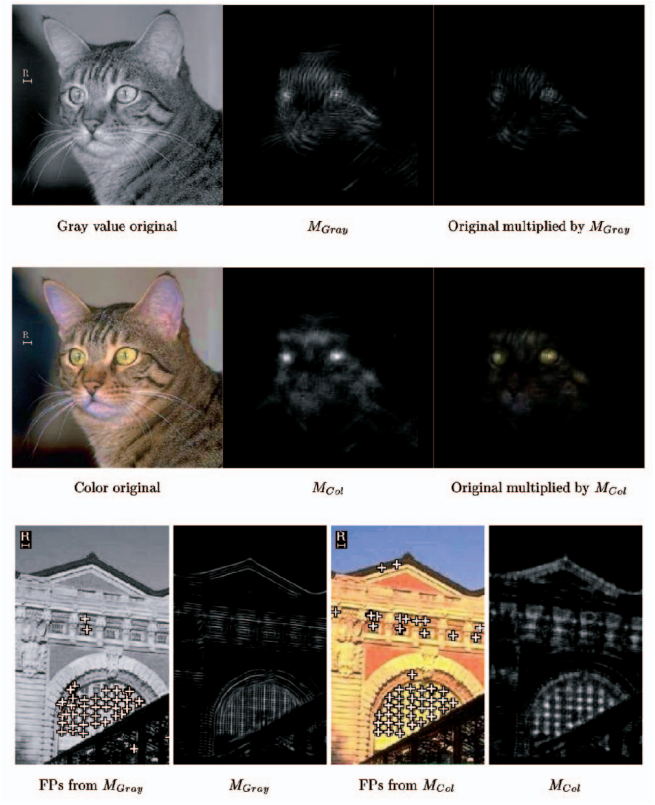

Fig. 4. Exploitation of color is most fruitful when symmetric regions differ from the surroundings more in color than in gray-value contrast. Maxima detection of the color-based symmetry map is more stable (cats' eyes). Bottom row: Due to the poor quality of the original, M_{Gray} detects only the windows under the arc which have white frames around dark glass. In M_{Col}, additionally, the symmetry of the upper row of windows stands out. The 50 best FPs are displayed. Note not all FP-markers are visible due to very close maxima, in particular, for M_{Gray}, maxima are crowded under the arc. Both images are from [48]: Cat: "domesticated461," building: "architecture1051."

4.2 Saliency Detectors for Comparison

Two other methods will be used for comparison: The detector of Harris and Stephens [21] and the Daubechies4 transform [14].

The calculation of the saliency map M_{Harris} after Harris and Stephens [21], and the choice of parameters are outlined in the Appendix. FPs are detected at the highest maxima of a smoothed version of M_{Harris}; this method will be referred to as **Harris**. Harris was chosen for comparison because it could be shown superior to other, related approaches [12], [16], [27], [28] in the work of Schmid et al. [57]. Interest points obtained from M_{Harris} proved to be the most stable against 2D image rotations, scaling, lighting variation, viewpoint change, and camera noise. Moreover, Harris produced interest points with a higher information content in terms of the entropy measure proposed in [57].

The algorithm referred to as **Daubechies4** is based on wavelet filtering. SPs detected by a multiresolution wavelets approach turned out to cohere with human eye fixations in the investigations of Privitera and Stark [50]. Here, the SP-detector proposed by Tian et al. is adopted, which proved to be well-suited for the representation of local image features for context-based image retrieval [62]. Because of its complexity, the approach cannot by outlined here, see [62] for details. In short, the image is convoluted with orthogonal Daubechies4 wavelets [14] at different resolutions which yield a complete and nonredundant representation of wavelet coefficients. Since this type of wavelet has a compact support, the set of pixels from which a certain coefficient was computed is known. Therefore, SPs can be found by a recursive search for high coefficients from coarse-to-fine resolutions. The algorithm is made available in the Internet at http://telesun. insa-lyon.fr/~loupias/points/demo.html. For a survey of wavelet theory, see, e.g., [41].

4.3 Repeatability in the Presence of Noise

Stability against signal noise was checked using 10 static indoor scenes for each of which two frames, F_1 and F_2, were recorded by a low-cost color CCD-camera. The average difference between corresponding pixels of F_1 and F_2 was 2.1 for the red channel, 2.3 for green, and 5.8 for blue (256 levels for each channel).

The *ε-repeatability rate* $r(\varepsilon)$ is the fraction of FPs detected in F_1 which can be found in F_2 within an ε-neighborhood of the original position:

$$r(\varepsilon) = \frac{\left|\{p_i^1 \text{ found in } F_1 \mid \exists\, p_j^2 \text{ found in } F_2 \wedge \|p_i^1 - p_j^2\| \leq \varepsilon\}\right|}{N_{FP}} \quad \text{with } i,j = 1 \ldots N_{FP}, \tag{19}$$

where N_{FP} is the number of FPs found in F_1. Using the best $N_{FP} = 10$ FPs detected by each algorithm, all algorithms reach good repeatability rates $r(\varepsilon = 1.5)$: ColSym 0.92, GraySym 0.93, Harris 0.96, and Daubechies4 0.92. Radius $\varepsilon = 1.5$ was chosen because it encircles the 8-neighborhood of the pixel in F_1.

Additional corruption of 5 percent of the pixels of F_2 with salt-and-pepper noise leads to reduced repeatability rates $r(\varepsilon = 1.5)$: ColSym 0.82, GraySym 0.79, Harris 0.72, and Daubechies4 0.54.

4.4 Stability Against 3D Object Rotation

According to the criteria defined in Section 1.3, FPs should be stable against object rotations over a certain angular range. But, what is "stability"? "Repeatability" as defined above is not applicable for 3D object rotation because repeatability requires a unique relation between object points of different images. For planar rotation, this relation is a simple homography. Therefore, in [57], SP-stability is tested by rotation of the camera around its optical axis.

In the more realistic case of an object rotating in 3D, the problem is more complex than finding a point-to-point relation for different frames. Consider an object in a reference pose with an FP at $p \in \mathbb{R}^2$ in image coordinates which corresponds to a point on the surface of the object with world coordinates $S \in \mathbb{R}^3$. If the object is rotated to another pose, the surface point changes its world coordinates to S'. The "naive" definition of FP-stability would be that the FP must be found at the projection of S' to the image plane, p'. This definition, however, is unsatisfactory since p' does not necessarily have the same "saliency properties" which made p an FP. For example, let's consider the "degenerate case" of a

Fig. 5. Gray-value contrast is low in the "Apples" and "Flowers" image, so only ColSym detects the obvious symmetry of the apples and the flowers. The Lena image offers good gray-value contrast, so both GraySym and ColSym are successful. FPs are located at the highest maxima of the symmetry maps. Images: Apples: "food303" of [48], flowers: "flowers3762" of [48].

perfect sphere with an FP p in its center, detected because of symmetry. Rotation of the sphere shifts the original location S *on the sphere* to S', but the FP will stay in the center (at p)—which is perfectly all right both from a semantical point-of-view and from the signal properties because the symmetric sphere looks the same as before. Nevertheless, the naive definition of FP-stability is hurt.

Consequently, a stability rating must include the semantics of FPs. As a database, the Columbia Object Image Library (COIL-100) [47], [45] was used. The library consists of RGB-images of objects presented on a turntable that is rotated by 360 degrees in steps of 5 degrees, so there are 72 images of each object. An FP is considered stable if it appears over at least three subsequent frames at the same part of the object, the idea being that there must be at least one frame for which the FP is found also for $\pm 5^\circ$ rotations. Stronger stability criteria, e.g., presence of the FP for $\pm 10^\circ$, are reasonable when a method must be judged for a particular application. However, such judgments can be easily derived once the FPs have been tracked using the $\pm 5^\circ$-criterion. The key problem is that a human must decide whether an FP "stays in place" or not, so the result depends on human interpretation.

Fig. 6 shows an example from COIL-100 together with a recording of FP-stability. The best six FPs of ColSym are tracked over 24 frames from 0 degrees to 115 degrees. A green line indicates a stable FP that must range over at least three frames. In this case, the FP is counted a *true positive* in each frame. A red line indicates an unstable FP, ranging over one or two frames. The FP is then counted a *false positive* in each frame. In some cases, green lines are interrupted by dashed red lines, meaning one or two missing links in a chain of an otherwise stable FP. Such missing FPs are counted as *false negatives*.

Some FPs in Fig. 6 can be easily judged as stable, e.g., the FP on the left eye is stable over six frames (frame 1 to 6). The FP on the right foot is stable from frame 1 to 24 with interruptions at frame 2 and 18, 19. In total, there are $N_{tp} = 130$ true positives, $N_{fp} = 14$ false positives, and $N_{fn} = 5$ false negatives.

The stability measure Q is defined as:

$$Q = 1 - \frac{\#\text{ wrongly placed FPs}}{\#\text{ all FPs}} = 1 - \frac{N_{fp} + N_{fn} - N_{CO}}{N_{tp} + N_{fp}}. \tag{20}$$

N_{CO} denotes the number of cases where false negatives ("missing links") coincide with the appearance of false

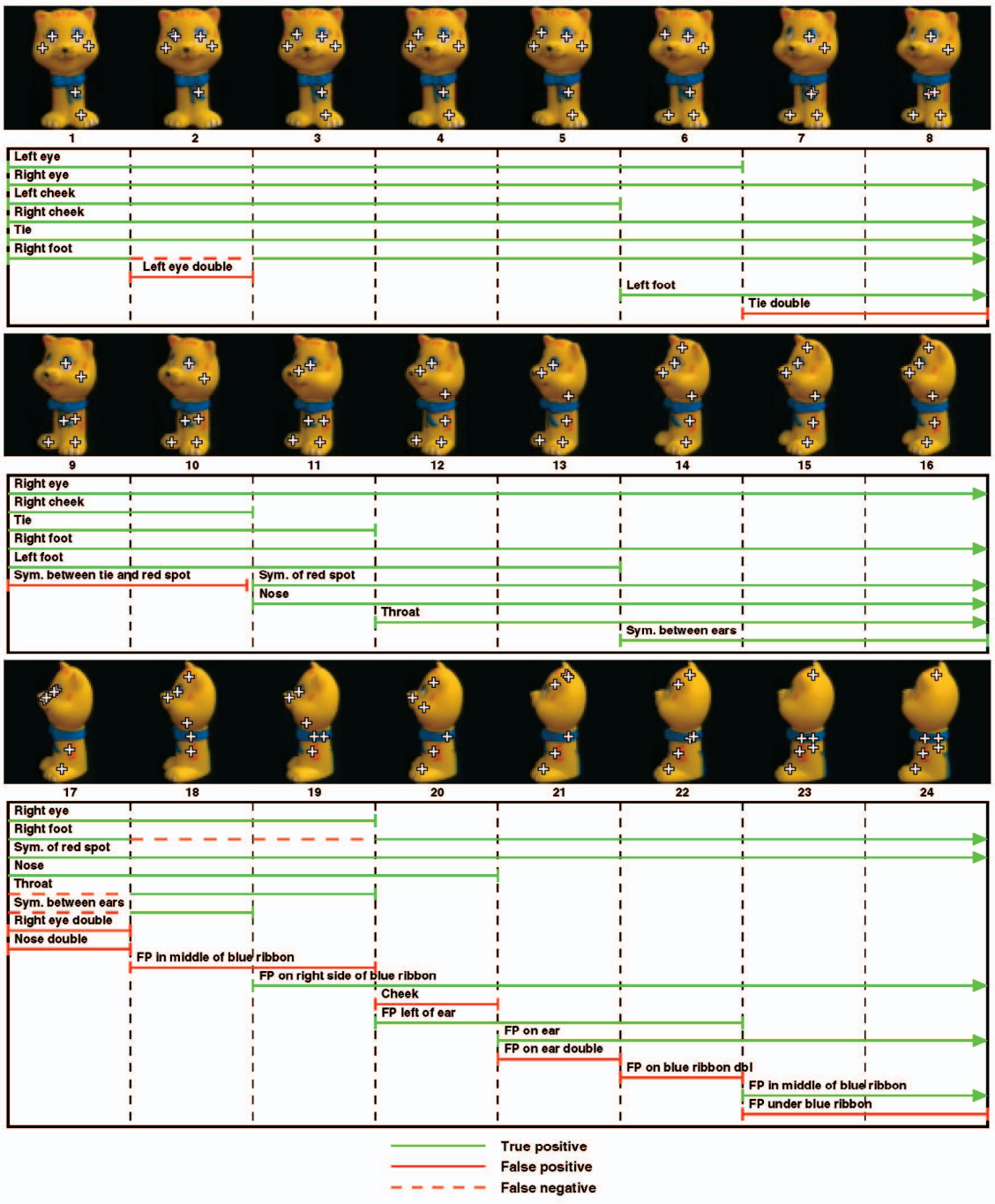

Fig. 6. Testing FP-stability for 3D object rotation: An FP is counted as "stable" if it is located at the same part of the object over at least three subsequent frames (true positives). Unstable FPs appearing only in one or two connected frames are false positives. An interruption of no more than two frames in an otherwise stable chain of FPs is counted as false negative. $N_{FP} = 6$ is fixed for each frame. Image series: "obj17" of [47].

positives elsewhere, i.e., an FP temporarily shifted to another location. N_{CO} is subtracted to avoid *two* errors being counted in this case, so $Q \geq 0$ holds. This procedure is justified by the way FPs are generated from maxima of a continuous saliency map: Since only the N_{FP} highest maxima are accepted, the N_{FP}th and $(N_{FP}+1)$th maxima may easily change positions in the height ranking, which leads to a "hopping" FP. Subtraction of N_{CO} avoids such errors being counted twice. In case all errors are due to hopping FPs, $N_{fp} = N_{fn} = N_{CO}$, so Q reduces to the percentage of true positives of all FPs: $Q = N_{tp}/(N_{tp} + N_{fp})$.

The problem of FP-tracking being subject to human judgment is illustrated in Fig. 6, e.g., by the FP which indicates, at first (frames 9 and 10), the symmetry between the blue tie and the red spot on the body of the figure. It moves continuously toward the red spot and remains there from frame 11 to 24. Here, the FP was considered false positive in frames 9 and 10 because it shifts from a yellow to a red spot. However, frames 9 and 10 could also be interpreted as stable because the shift of the FP is a continuous movement.

FPs were generated for eight different objects of COIL-100 over the entire 360°-range. The original resolution of

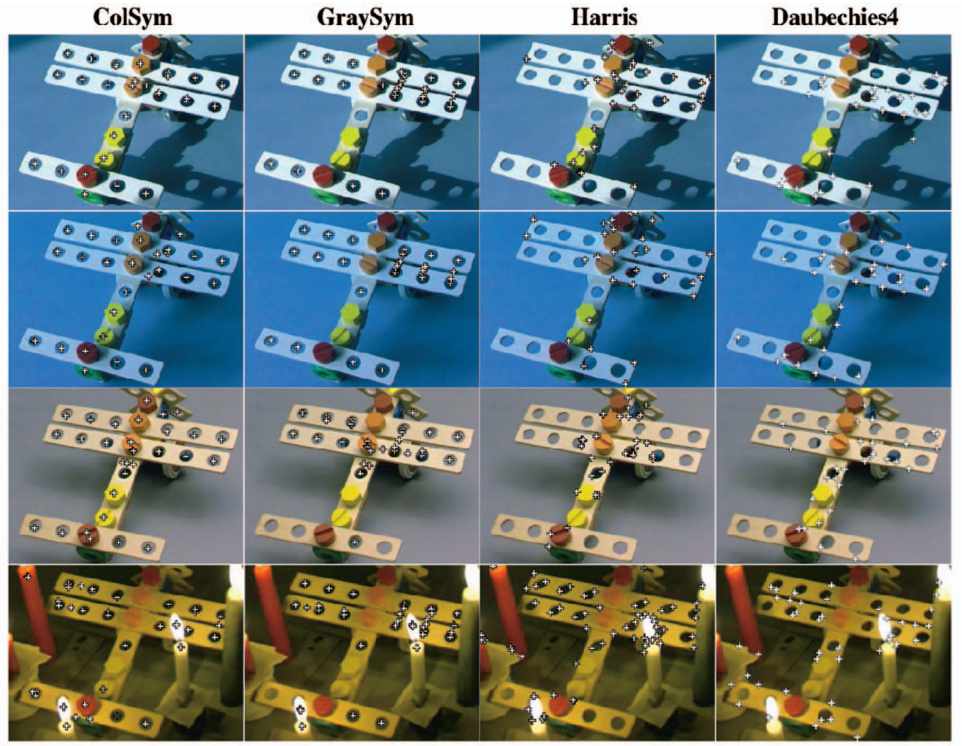

Fig. 7. Top to bottom: Direct sunlight from a window, diffuse daylight from a window on a cloudy day, diffused halogen lighting, five candles. To increase difficulty, the camera color temperature was fixed to optimal for sunlight, so the cloudy daylight images appear bluish. The symmetry-based algorithms both find holes, but the obvious symmetry of the colored bolt heads is detected only by ColSym. Harris and Daubechies4 detect corners and edges, most of which do not facilitate recognition of the assembled objects.

128×128 was subsampled to 64×64. All FP-detectors reached good results:

$$Q_{ColSym} = 91.6\%, \quad Q_{GraySym} = 90.3\%,$$
$$Q_{Harris} = 94.0\%, \quad Q_{Daubechies4} = 84.1\%.$$

4.5 Repeatability for Varying Lighting

FP-stability against variations of lighting was tested using the repeatability rate $r(\varepsilon)$, (19). Images were acquired from the "Baufix®-scenario" of the collaborative research project "Situated Artificial Communicators" at Bielefeld University, Germany (see, e.g., [54], [7], [5], [61]). The objects are wooden toy pieces of the Baufix®—system. These objects are well-suited for a comparison of different types of algorithms because they offer symmetries as well as corners and edges. Fig. 7 shows an example for four different lighting conditions (top to bottom): direct sunlight, diffuse daylight on a cloudy day, diffused halogen lamps, and light from five candles. For all four lighting conditions, images of 10 different Baufix-scenes, including single and assembled pieces, were recorded.

For each algorithm, the $N_{best} = 30$ FPs were evaluated. Since the candles attract some FPs, N_{best} was increased by this number individually for each algorithm in the last experiment. The symmetry-based algorithms both detect the holes, but only ColSym is able to also find the symmetries of the colored bolt heads. Harris and Daubechies4 concentrate on corners and edges. Since none of the lighting conditions are predetermined as a reference F_1 (in terms of Section 4.3), the repeatability rate $r(\varepsilon = 3)$ was averaged over the 12 pairs of different lighting conditions (e.g., candles as F_1 and sunlight as F_2). The tolerance of $\varepsilon = 3$ is chosen larger than in Section 4.3 because the original scenes were built in sunlight and had to be reconstructed on a cloudy day, which leads to slight differences. The averaged repeatability rates are: ColSym 0.45, GraySym 0.48, Harris 0.18, and Daubechies4 0.14.

The superiority of the symmetry-based methods is obvious and can partly be explained by the fact that long edges do not offer outstanding points, so even small changes in lighting make FPs move along edges. Most important, however, for an actual recognition system is that ColSym produces FPs on semantically meaningful visual entities, as becomes clear in Section 5.

4.6 Judging Saliency

Finding a measure for the "saliency" or "distinctiveness" of a point is the most difficult task. Schmid et al. [57] compute rotation invariants from a local jet of Gaussian derivatives [34] as descriptors for local gray-value patterns around SPs. They argue that most information is conveyed within a population of observed SPs if the descriptors are not similar to each other but spread out, which can be measured in terms of entropy.

However, this method has a serious drawback: Calculation of entropy requires that the descriptor space is partitioned into cells, then the number of descriptors within the cells is counted. Because of the curse of dimensionality, this is possible only for a low-dimensional descriptor space—else an enormous number of sample SPs and descriptors would be needed. Therefore, in [57], only a four-dimensional descriptor space is used, which leads to a poor representation of shape around an SP in, e.g., a 20×20-window.

Therefore, another method is used here to measure distinctiveness of an observed FP population: The principal components (PCs) of FP-centered windows are compared to those centered at random points. Hancock et al. extracted the "natural" PCs from gray-level images, i.e., the PCs of a collection of randomly chosen image patches [18]. They used a single layer feed forward neural network proposed by Sanger [55] for the successive calculation of the PCs with the largest eigenvalues. The d-dimensional training vectors $\vec{x} \in \mathbb{R}^d$ represent the gray values of square image patches of $d = w \times w$ pixels from which the mean gray-value over all images was subtracted. The net has one node for each PC with an input weight vector $\vec{W}_i \in \mathbb{R}^d$ with $i = 1 \ldots n$ for n different PCs. The activation V_i of the nodes is calculated using the linear function

$$V_i \;=\; \sum_{j=1}^{d} W_{ij} x_j, \quad i = 1 \ldots n. \tag{21}$$

After training by Sanger's adaptation rule,

$$\Delta W_{ij} = \epsilon V_i \left[\left(x_j - \sum_{k=1}^{i-1} V_k W_{kj} \right) - V_i W_{ij} \right], \quad i = 1 \ldots n, \tag{22}$$

the weight vectors represent the PCs in the order of the corresponding eigenvalues, beginning with the largest.

To evaluate the FP-detectors, the 12 PCs with the highest eigenvalues were computed after (21) and (22) for randomly chosen points ("natural PCs") and for windows centered by the four FP-algorithms. The window size was 63×63, each window was masked by a Gaussian with 10 pixels standard deviation to avoid edge effects. As a database, the *photos*-section of [48] was used which contains over 89,000 photographs covering a large range of topics like landscapes, cityscapes, people, animals, entertainment, or sports to name but a few. Convergence was achieved with a learning rate ϵ decreasing exponentially from 1.0 to 0.01 over $5 \cdot 10^6$ adaptation steps. Images were chosen at random from the collection, then 10 points were chosen either randomly for the natural PCs or by the FP-detector for FP-centered PCs. To ensure orthonormality of the obtained weight vectors, Gram-Schmidt orthonormalization [29] was carried out after the training. This has no visible effect but makes the difference estimation outlined below possible.

Fig. 8 clearly shows that the PCs of FP-centered windows differ from the natural PCs. To quantify this difference, let the orthonormalized natural PCs of the n largest eigenvalues be denoted by $\vec{u}_1 \ldots \vec{u}_n$ and the orthonormalized PCs of FP-centered windows by $\vec{v}_1 \ldots \vec{v}_n$. Note that

$$\vec{u}_i \cdot \vec{u}_j = \delta_{ij} \quad \text{and} \quad \vec{v}_i \cdot \vec{v}_j = \delta_{ij} \quad \text{for} \quad i, j = 1 \ldots n. \tag{23}$$

The most simple approach to compare the sets $\{\vec{u}_1 \ldots \vec{u}_n\}$ and $\{\vec{v}_a \ldots \vec{v}_n\}$ would be to check how many of the vectors $\vec{v}_i$ are within $\mathrm{span}\{\vec{u}_1 \ldots \vec{u}_n\}$, i.e., which $\vec{v}_i$ are linear combinations of the $\{\vec{u}_1 \ldots \vec{u}_n\}$. This method, however, would be inappropriate: Consider the first, Gaussian-like natural PC $\vec{u}_1$. Even if $\vec{v}_1$ was equal to $\vec{u}_1$ for all but one pixel, $\vec{v}_1$ would probably *not* be within $\mathrm{span}\{\vec{u}_1 \ldots \vec{u}_n\}$ as long as $n \ll d$, because the PCs with high eigenvalues mostly represent low spatial frequencies. So, a high-frequency perturbation like the single disrupted pixel of $\vec{v}_1$ cannot be represented by $\{\vec{u}_1 \ldots \vec{u}_n\}$.

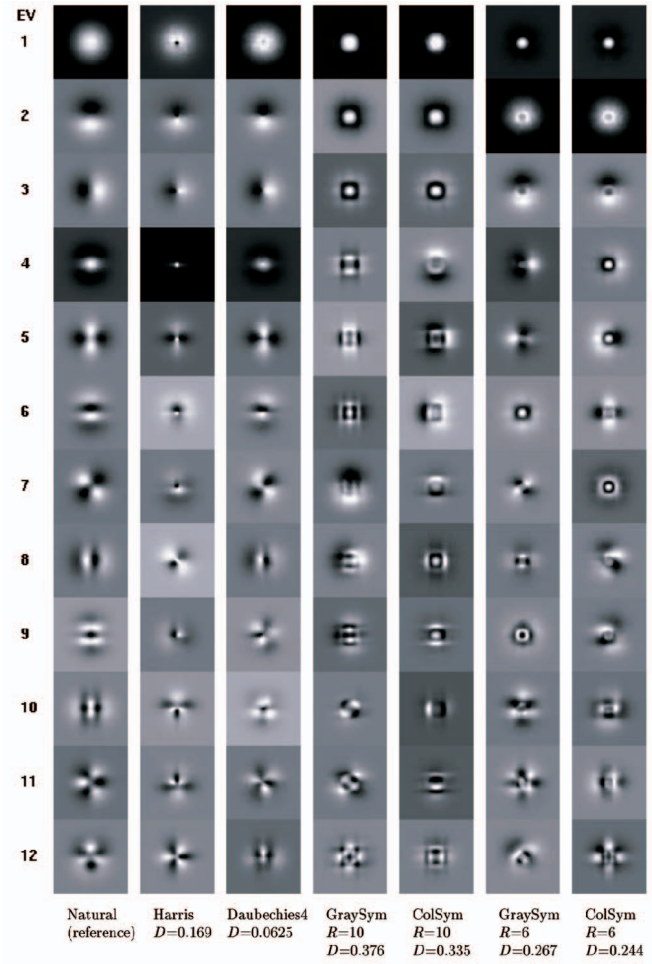

Fig. 8. PCs of randomly chosen windows of natural images (left) and of windows chosen by the FP-detectors. Note the algorithm of Sanger leads to PCs with random signs, so signs of the PCs were chosen to facilitate visual comparison.

A better measure D for the "difference" between $\{\vec{u}_1 \ldots \vec{u}_n\}$ and $\{\vec{v}_1 \ldots \vec{v}_n\}$ is the mean square reconstruction error of the $\vec{u}_i$ in the basis $\{\vec{v}_1 \ldots \vec{v}_n\}$:

$$D \;=\; \frac{1}{n} \cdot \sum_{i=1}^{n} \left\| \vec{u}_i - \sum_{j=1}^{n} a_{ij} \vec{v}_j \right\|^2 \qquad \text{with} \quad a_{ij} = \vec{u}_i \cdot \vec{v}_j. \tag{24}$$

For the evaluation, the $n = 12$ PCs with the largest eigenvalues were used since they capture most of the variance within the windows and can be viewed as a representation of the "typical" surroundings of the FPs. Using the natural PCs as reference, the following differences D were computed for the FP-centered PCs:

$$\begin{array}{ll} \mathrm{ColSym}(R = 10)\ 0.335, & \mathrm{GraySym}(R = 10)\ 0.376, \\ \mathrm{ColSym}(R = 6)\ 0.244, & \mathrm{GraySym}(R = 6)\ 0.267, \\ \mathrm{Harris}\ 0.169, & \mathrm{Daubechies4}\ 0.0625. \end{array}$$

A major advantage of the PC-based saliency judgment is that visualization of the PCs facilitates interpretation considerably. The smallest difference D to natural PCs was found for Daubechies4. The reason becomes obvious in

Fig. 8: Most of the PCs resemble the natural ones; differences become visible only for higher spatial frequencies (note spatial frequencies of the PCs roughly increase with decreasing eigenvalues). This behavior is not surprising as FPs from Daubechies4 are located mostly on edges.

For Harris, the detector with second smallest D, even the first PCs differ strongly from the natural PCs—but only in the center of the window, which indicates the large amount of focused corners. Apart from the center, the first three PCs are almost equal to the natural PCs and, also, the PCs of higher spatial frequencies resemble the natural ones, though slightly out of order. A likely explanation for these effects is that Harris works on a small scale and detects only relatively simple structures. Windows detected by Harris mainly have corners at the center, so this part is guaranteed to be different from the average window. But, since Harris does not evaluate the area *around* corners, the rest of the focused windows are not necessarily different from random windows. In other words, corners are "special" only on the small scale detected by Harris, but not on a larger scale—the surroundings of corners look like other parts of an image, at least from a statistical point of view.

The focused symmetric image windows are more "special" than corners and edges in terms of (24). Fig. 8 shows why D is much higher for the symmetry-based algorithms: PCs of both high *and* low-spatial frequencies differ considerably from the natural PCs. Some of the PCs clearly reflect the symmetry of the focused image windows. A larger symmetry radius R increases this effect because it ensures that a larger part of the focused windows is symmetric which increases the "statistical difference" to random windows.

5 Sample Application

The major benefit of ColSym is that FPs can be detected even for low gray-value contrast. In the Baufix-scenario of the collaborative research project SFB 360 at Bielefeld University, detection of toy pieces within assemblies is required in the context of a human-machine interaction scenario, where a robot assembles the toy pieces guided by hand gestures and speech from a human instructor. For details see, e.g., [54], [7], [5], [61]. As many of the toy pieces have similar gray values, color is necessary: Fig. 7 shows that the symmetric colored bolt heads can be detected only by ColSym, not by GraySym.

To verify that ColSym indeed simplifies the task of view-based object recognition as outlined in Section 1.2, a simple classification system was trained to recognize four classes of parts of the objects which are relevant for the robot system: Holes of bars, holes of nuts, holes of cubes, and heads of bolts. FPs were generated and hand labeled on 80 sample images. From the FP-centered windows, 20 PCs were calculated. As pointed out in Section 1.1, the window size is application specific; here it is approximately the size of a bolt head. Projections of the sample windows onto the PCs were then used to establish a simple nearest neighbor classifier C_1. C_1 reaches a rate of 89 percent correctly classified FPs on 30 test images, including even difficult scenes, as shown in Fig. 9. According to the requirements pointed out in Section 1.3, most but not all FPs are well-positioned on objects. Ill-positioned FPs were assigned to an additional rejection class. The classifier puts no label on the FP in this case.

In a second experiment, the FP-detection was replaced by random point generation. Of the random points, only those were used which are on objects (not on background), however, *no particular position* on the object was required. From these windows, 20 PCs were calculated, then a nearest neighbor classifier C_2 was established from the (labeled) window projections. On test samples, C_2 yields only 23 percent correct classifications because the translatory gitter of unfocused objects makes appearance much more complex on the signal level.

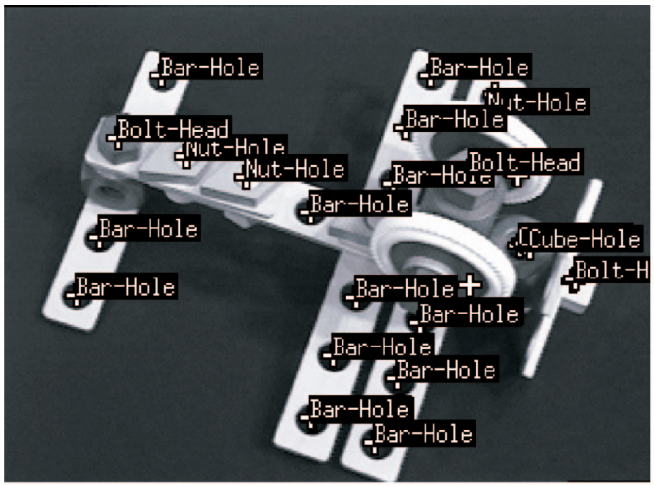

Fig. 9. Application of ColSym in the framework of an object recognition system. There are four classes: Holes in wooden bars, cubes, and rhomb-shaped nuts plus bolt heads. The class "bolt head" has the highest complexity, comprised of two different shapes and four different colors. The exact location of the FPs allows good recognition rates even for complex scenes. FPs classified as "unknown" have no label.

This simple view-based recognition system was chosen to demonstrate the use of stable FPs in a way easy to reproduce. For the actual robot scenario, a much more elaborate vision system is used which is based on several modules for FP-generation (including ColSym), subsequent feature extraction by local PCA and feature classification by several neural networks. For details, see [22], [7], [26], [23], [25].

6 Summary and Conclusion

In this paper, the role of focus points (FPs) for view based object recognition was outlined. The key features FPs must have are stability, distinctiveness, and being located on parts of objects which are relevant for recognition.

As a means to detect FPs, the algorithm ColSym was proposed which is based on the context-free symmetry measure (GraySym) used by Reisfeld et al. [53]. ColSym exploits color to detect symmetries even when they are not visible from gray-value contrast alone. It was shown that other extensions of GraySym to color, which are simpler than ColSym, do not work.

Together with two other algorithms, Harris and Daubechies4, both symmetry-based methods were evaluated for the three criteria set up for FPs. All algorithms provide good stability against noise and object rotation, but ColSym and GraySym are considerably better for varying lighting. To judge FP-distinctiveness free of context, the principal components of FP-centered image windows were calculated and compared to those of random windows. It turns out that image windows focused by ColSym and GraySym are, on average, more distinctive than corners and edges.

While the performance of ColSym is comparable to GraySym in stability and distinctiveness, the major advantage of ColSym is its usability: Stable FPs can be generated on many semantically meaningful visual entities, in particular, on colored objects which cannot be detected using GraySym. This makes ColSym a useful module in the view-based object recognition system outlined shortly in Section 5. So, the initial demand that FP-detectors must prove applicability in a real vision system could be fulfilled. The examples show that color symmetries coincide with meaningful objects not only in the scenario of toy objects, but also in real world images. Naturally, symmetry-based FP-detection cannot guide visual attention stand alone, so the cooperation with other saliency features will be the topic of future investigation.

APPENDIX

THE DETECTOR OF HARRIS AND STEPHENS

The corner and edge detector proposed by Harris and Stephens [21] is based on first derivatives of the signal I_x, I_y (see (1)) from which a matrix A is calculated:

$$A(p) \;=\; \begin{pmatrix} \langle I_x^2 \rangle_{W(p)} & \langle I_x I_y \rangle_{W(p)} \\ \langle I_x I_y \rangle_{W(p)} & \langle I_y^2 \rangle_{W(p)} \end{pmatrix}. \qquad (25)$$

Here, $\langle\,\rangle_{W(p)}$ denotes a weighted averaging over a window $W(p)$ centered at p. The weight function inside the window is a Gaussian. A is an approximation of the autocorrelation function of the signal, see, e.g., [57]. An SP is detected if both eigenvalues of A are large. To reduce the computational effort, the saliency map is calculated from

$$M_{Harris}(p) \;=\; \det(A) - \alpha \cdot (\mathrm{Trace}(A))^2. \qquad (26)$$

In the implementation used here, derivatives I_x, I_y are computed by 5×5-Sobel operators. A Gaussian with $\sigma = 2$ is used to weight the components of A within window W. As suggested in [57], a value of 0.06 is used for the constant α in (26).

ACKNOWLEDGMENTS

The author would like to thank Helge Ritter, head of the Neuroinformatics Department at Bielefeld University, for the long term support of his work. This work was funded by the Deutsche Forschungsgemeinschaft (DFG) within the collaborative research project SFB 360 "Situated Artificial Communicators."

REFERENCES

[1] J.Y. Aloimonos, I. Weiss, and A. Bandyopadhyay, "Active Vision," *Int'l J. Computer Vision,* vol. 1, pp. 334-356, 1987.

[2] H. Asada and M. Brady, "The Curvature Primal Sketch," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 8, no. 1, pp. 2-14, 1986.

[3] G. Backer, B. Mertsching, and M. Bollmann, "Data and Model-Driven Gaze Control for an Active-Vision System," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 12, pp. 1415-1429, Dec. 2001.

[4] R. Bajcsy and M. Campos, "Active and Exploratory Perception," *Computer Vision, Graphics, and Image Processing: Image Understanding,* vol. 56, no. 1, pp. 31-40, 1992.

[5] C. Bauckhage, G.A. Fink, G. Heidemann, N. Jungclaus, F. Kummert, S. Posch, H. Ritter, G. Sagerer, and D. Schlüter, "Towards an Image Understanding Architecture for a Situated Artificial Communicator," *Pattern Recognition and Image Analysis,* vol. 9, no. 4, pp. 542-550, 1999.

[6] P. R. Beaudet, "Rotationally Invariant Image Operators," *Proc. Fourth Int'l Joint Conf. Pattern Recognition,* pp. 579-583, 1978.

[7] E. Braun, G. Heidemann, H. Ritter, and G. Sagerer, "A Multi-Directional Multiple Path Recognition Scheme for Complex Objects," *Pattern Recognition Letters,* special issue on Pattern Recognition in Practice VI, June 1999.

[8] S. Bres and J.M. Jolion, "Detection of Interest Points for Image Indexation," *Proc. Third Int'l Conf. Visual Information Systems, Visual '99,* pp. 424-434, 1999.

[9] V. Bruce and M. Morgan, "Violation of Symmetry and Repetition in Visual Pathways," *Perception,* vol. 4, pp. 239-249, 1975.

[10] K. Brunnström, J.-O. Eklundh, and T. Uhlin, "Active Fixation for Scene Exploration," *Int'l J. Computer Vision,* vol. 17, pp. 137-162, 1996.

[11] C.-H. Chen, J.-S. Lee, and Y.-N. Sun, "Wavelet Transformation for Gray-Level Corner Detection," *Pattern Recognition,* vol. 28, no. 6, pp. 853-861, 1995.

[12] J.C. Cottier, "Extraction et Appariements Robustes des Points d'Intért de Deux Images non Étalonnées," technical report, LIFIA-IMAG-INRIA, Rhone-Alpes, 1994.

[13] J.L. Crowley, P. Bobet, and M. Mesrabi, "Camera Control for an Active Camera Head," *Pattern Recognition and Artificial Intelligence,* J. L. Crowley, P. Stelmaszyk, T. Skordas, and P. Puget, vol. 7, no. 1, 1993.

[14] I. Daubechies, "Orthonormal Bases of Compactly Supported Wavelets," *Comm. Pure and Applied Math.,* vol. 41, pp. 909-996, 1988.

[15] L. Dreschler and H.-H. Nagel, "Volumetric Model and 3D Trajectory of a Moving Car Derived from Monocular TV Frame Sequences of a Street Scene," *Computer Graphics and Image Processing,* vol. 20, pp. 199-228, 1982.

[16] W. Förstner, "A Framework for Low Level Feature Extraction," *Proc. Third European Conf. Computer Vision,* pp. 383-394, 1994.

[17] W.T. Freeman and E.H. Adelson, "The Design and Use of Steerable Filters," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 13, pp. 891-906, 1991.

[18] P.J.B. Hancock, R.J. Baddeley, and L.S. Smith, "The Principal Components of Natural Images," *Network,* vol. 3, pp. 61-70, 1992.

[19] U. Handmann, T. Kalinke, C. Tzomakas, M. Werner, and W. v. Seelen, "An Image Processing System for Driver Assistance," *Image and Vision Computing,* vol. 18, no. 5, pp. 367-376, 2000.

[20] R.M. Haralick and L.G. Shapiro, *Computer and Robot Vision,* vol. 2, 1993.

[21] C. Harris and M. Stephens, "A Combined Corner and Edge Detector," *Proc. Fourth Alvey Vision Conf.,* pp. 147-151, 1988.

[22] G. Heidemann, "Ein Flexibel Einsetzbares Objekterkennungssystem auf der Basis Neuronaler Netze," PhD thesis, Infix, DISKI 190, Univ. Bielefeld, Technische Fakultät, 1998.

[23] G. Heidemann, D. Lücke, and H. Ritter, "A System for Various Visual Classification Tasks Based on Neural Networks," *Proc. 15th Int'l Conf. Pattern Recognition,* A. Sanfeliu et al., eds., vol. I, pp. 9-12, 2000.

[24] G. Heidemann, T.W. Nattkemper, and H. Ritter, "Farbe und Symmetrie für die datengetriebene Generierung prägnanter Fokuspunkte," *Proc. Fourth Workshop Farbbildverarbeitung,* V. Rehrmann, ed., pp. 65-71, 1998.

[25] G. Heidemann and H. Ritter, "Combining Multiple Neural Nets for Visual Feature Selection and Classification," *Proc. Ninth Int'l Conf. Artificial Neural Networks,* pp. 365-370, 1999.

[26] G. Heidemann and H. Ritter, "Efficient Vector Quantization Using the WTA-Rule with Activity Equalization," *Neural Processing Letters,* vol. 13, no. 1, pp. 17-30, 2001.

[27] F. Heitger, L. Rosenthaler, R. von der Heydt, E. Peterhans, and O. Kübler, "Simulation of Neural Contour Mechanism: From Simple to End-Stopped Cells," *Vision Research,* vol. 32, no. 5, pp. 963-981, 1992.

[28] R. Horaud, F. Veillon, and T. Skordas, "Finding Geometric and Relational Structures in an Image," *Proc. First European Conf. Computer Vision,* pp. 374-384, 1990.

[29] A.S. Householder, *The Theory of Matrices in Numerical Analysis.* New York: Dover Publications, 1964.

[30] L. Itti, C. Koch, and E. Niebur, "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 20, no. 11, pp. 1254-1259, Nov. 1998.

[31] T. Kalinke and W. von Seelen, "Entropie als Maß des Lokalen Informationsgehalts in Bildern zur Realisierung einer Aufmerksamkeitssteuerung," *Mustererkennung,* B. Jähne, P. Geißler, H. Haußecker, and F. Hering, eds., pp. 627-634, Heidelberg: Springer Verlag, 1996.
[32] L. Kaufman and W. Richards, "Spontaneous Fixation Tendencies for Visual Forms," *Perception and Psychophysics,* vol. 5, no. 2, pp. 85-88, 1969.
[33] L. Kitchen and A. Rosenfeld, "Gray-Level Corner Detection," *Pattern Recognition Letters,* vol. 1, pp. 95-102, 1982.
[34] J.J. Koenderink and A.J. van Doorn, "Representation of Local Geometry in the Visual System," *Biological Cybernetics,* vol. 55, pp. 367-375, 1987.
[35] R. Laganiére, "Morphological Corner Detection," *Proc. Sixth Int'l Conf. Computer Vision,* pp. 280-285, 1998.
[36] J.-S. Lee, Y.-N. Sun, and C.-H. Chen, "Multiscale Corner Detection by Using Wavelet Transform," *IEEE Trans. Image Processing,* vol. 4, no. 1, pp. 100-104, 1995.
[37] T. Lindeberg, "Detecting Salient Blob-Like Image Structures and Their Scales with a Scale-Space Primal Sketch: A Method for Focus-of-Attention," *Int'l J. Computer Vision,* vol. 11, no. 3, pp. 283-318, 1993.
[38] T. Lindeberg, "Scale-Space Theory: A Basic Tool for Analysing Structures at Different Scales," *J. Applied Statistics,* vol. 21, no. 2, pp. 224-270, 1994.
[39] P.J. Locher and C.F. Nodine, "Symmetry Catches the Eye," *Eye Movements: From Physiology to Cognition,* A. Levy-Schoen and J.K. O'Reagan, eds., pp. 353-361, B.V. (North Holland): Elsevier Science, 1987.
[40] J. Maintz and M. Viergever, "A Survey of Medical Image Registration," *Medical Image Analysis,* vol. 2, no. 1, pp. 1-36, 1998.
[41] S. Mallat, *A Wavelet Tour of Signal Processing.* Academic Press, 1998.
[42] G. Medioni and Y. Yasumoto, "Corner Detection and Curve Representation Using Cubic B-Splines," *Computer Vision, Graphics, and Image Processing,* vol. 39, pp. 267-278, 1987.
[43] B. Moghaddam and A. Pentland, "Probabilistic Visual Learning for Object Representation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, pp. 696-710, July 1997.
[44] H.P. Moravec, "Towards Automatic Visual Obstacle Avoidance," *Proc. Fifth Int'l Joint Conf. Artificial Intelligence,* pp. 584-587, 1977.
[45] H. Murase and S.K. Nayar, "Visual Learning and Recognition of 3-D Objects from Appearance," *Int'l J. Computer Vision,* vol. 14, pp. 5-24, 1995.
[46] T.W. Nattkemper, "Untersuchung und Erweiterung eines Ansatzes zur modellfreien Aufmerksamkeitssteuerung durch lokale Symmetrien in einem Computer Vision System," master's thesis, Technische Fakultät, Bielefeld Univ., Jan. 1997.
[47] S.A. Nene, S.K. Nayar, and H. Murase, "Columbia Object Image Library: COIL-100," Technical Report CUCS-006-96, Dept. of Computer Science, Columbia Univ., 1996.
[48] Art Explosion® Photo Gallery, Nova Development Corp., year?
[49] S. Palmer, "The Psychology of Perceptual Organization: A Transformational Approach," *Human and Machine Vision,* J. Beck, B. Hope, and A. Rosenfeld, eds., Academic Press, 1983.
[50] C.M. Privitera and L.W. Stark, "Algorithms for Defining Visual Regions-of-Interest: Comparison with Eye Fixations," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 9, pp. 970-982, Sept. 2000.
[51] R.P.N. Rao and D.H. Ballard, "An Active Vision Architecture Based on Iconic Representations," *Artificial Intelligence,* vol. 78, pp. 461-505, 1995.
[52] R.P.N. Rao, G.J. Zelinsky, M.M. Hayhoe, and D.H. Ballard, "Modeling Saccadic Targeting in Visual Search," *Advances in Neural Information Processing Systems 8,* D. Touretzky, M. Mozer, and M. Hasselmo, eds., MIT Press, 1995.
[53] D. Reisfeld, H. Wolfson, and Y. Yeshurun, "Context-Free Attentional Operators: The Generalized Symmetry Transform," *Int'l J. Computer Vision,* vol. 14, pp. 119-130, 1995.
[54] G. Rickheit and I. Wachsmuth, "Situated Artificial Communicators," *Artificial Intelligence Rev.,* vol. 10, pp. 165-170, 1996.
[55] T.D. Sanger, "Optimal Unsupervised Learning in a Single-Layer Linear Feedforward Neural Network," *Neural Networks,* vol. 2, pp. 459-473, 1989.
[56] C. Schmid and R. Mohr, "Local Grayvalue Invariants for Image Retrieval," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 5, pp. 530-535, May 1997.
[57] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of Interest Point Detectors," *Int'l J. Computer Vision,* vol. 37, no. 2, pp. 151-172, 2000.
[58] E. Shilat, M. Werman, and Y. Gdalyahu, "Ridge's Corner Detection and Correspondance," *Proc. Conf. Computer Vision and Pattern Recognition,* pp. 976-981, 1997.
[59] A. Shokoufandeh, I. Marsic, and S.J. Dickinson, "View-Based Object Recognition Using Saliency Maps," *Image and Vision Computing,* vol. 17, pp. 445-460, 1999.
[60] S. Smith and J. Brady, "SUSAN—A New Approach to Low Level Image Processing," *Int'l J. Computer Vision,* vol. 23, no. 1, pp. 45-78, 1997.
[61] J. Steil, G. Heidemann, J. Jockusch, R. Rae, N. Jungclaus, and H. Ritter, "Guiding Attention for Grasping Tasks by Gestural Instruction: The GRAVIS-Robot Architecture," *Proc. IEEE/RSJ Int'l Conf. Intelligent Robots and Systems,* 2001.
[62] Q. Tian, N. Sebe, M.S. Lew, E. Loupias, and T.S. Huang, "Image Retrieval Using Wavelet-Based Salient Points," *J. Electronic Imaging,* vol. 10, no. 4, pp. 835-849, 2001.
[63] C. Tomasi and T. Kanade, "Detection and Tracking of Point Features," Technical Report CMU-CS-91-132, Carnegie Mellon Univ., Pittsburgh, 1991.
[64] T. Tuytelaars and L. van Gool, "Content-Based Image Retrieval Based on Local Affinely Invariant Regions," *Proc. Third Int'l Conf. Visual Information Systems,* pp. 493-500, 1999.
[65] H. Zabrodsky, S. Peleg, and D. Avnir, "Symmetry as a Continous Feature," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 17, no. 12, pp. 1154-1166, Dec. 1995.
[66] Z. Zheng, H. Wang, and W. Teoh, "Analysis of Gray Level Corner Detection," *Pattern Recognition Letters,* vol. 20, pp. 149-162, 1999.
[67] B. Zitova, J. Kautsky, G. Peters, and J. Flusser, "Robust Detection of Significant Points in Multi-Frame Images," *Pattern Recognition Letters,* vol. 20, pp. 199-206, 1999.

Gunther Heidemann studied physics at the Universities of Karlsruhe and Münster and received the PhD degree in engineering from Bielefeld University in 1998. He is currently working within the collaborative research project "Hybrid Knowledge Representation" of the SFB 360 at Bielefeld University. His fields of research are mainly computer vision, robotics, neural networks, datamining, image retrieval, sonification, and hybrid systems.