A Comparison of Regional Feature Detectors in Panoramic Images

Kai Huebner University of Bremen Department of Mathematics and Computer Science D-28359 Bremen, Germany Email: khuebner@informatik.uni-bremen.de

Abstract— We present a novel approach to detect and describe visual features in panoramic image data. For various applications, especially computer and robot vision, robust and invariant features are key paths to explore scenes and objects. Most features applied in the literature can commonly be classified either as being local or being global. Local features characterize a significant point in the image like an edge. Global features describe a general property of the whole image like the color distribution. In this paper, we propose an in-between representation using region-based symmetry features. We compare the approach to a set of state-of-the-art affine feature detectors. Experiments show that the symmetry features are sparse, distinctive and robust to changes in panoramic image warp. Therefore, they are well applicable to robot vision tasks.

I. INTRODUCTION

Robust detection of features is a crucial task for applications that deal with visual information. Image data is highdimensional, complex and particularly sensitive to a multitude of changes. These mostly unpredictable changes greatly influence the image representation of one and the same object or scene. Therefore, a good feature detection is strongly required in dynamic and unrestricted real world environments. Preferably, this detection is invariant to a number of transformations like rotation, scaling, illumination change, perspective and panoramic warp, occlusion and image noise.

A visual feature is referred to as "good" if it seperates the core of information from the clutter. This basically depends on the application at hand and on the context it is used in. For our research on vision systems for mobile robots, we define a good feature to be both independent of transformations and distinctively repeatable in dynamic environments.

Most features applied in literature are commonly classified either as being *global*, *local* or *regional*. Concerning our definition of a good feature, common *global* features that describe general properties of an entire image scene are rather inappropriate for our task of robot scene interpretation. While single objects can be generalized easily by simple global features (e.g. size, color or texture attributes), finding stable and repeatable features is more complex for conglomerate scenes. However, global features give very compact representations of significant image properties. Therefore, global features are mainly used in image-based applications like image retrieval or image annotation. Daniel Westhoff, Jianwei Zhang University of Hamburg Technical Aspects of Multimodal Systems D-22527 Hamburg, Germany Email: {westhoff,zhang}@informatik.uni-hamburg.de

Many higher-level tasks like scene exploration or object classification and object tracking in complex scenes are grounded on *local* features. Being related to human visual perception, local visual features give clues for efficient scene exploration. They allow to focus on well-located interest points. Therefore, a variety of local features have been applied in a range of vision tasks, aiming at high robustness and repeatability. The Scale-Invariant Feature Transform (SIFT) proposed by Lowe [1] and the Harris-Laplacian by Mikolajczyk and Schmid [2] are two popular methods of local feature detection. While the SIFT uses local extrema of Differenceof-Gaussian (DoG) filters in scale-space to produce scaleinvariant features, the Harris-Laplace operator joins rotational invariant Harris features [3] and Laplacian scale-space analysis into an affine invariant interest point detector. As the exploration of invariant features is an active field of research, well elaborated comparisons of various feature detectors and descriptors have been published by Schmid et al. [4] and Mikolajczyk and Schmid [5], [6].

Due to the different characteristics of local and global features, it is beneficial for some applications to combine both approaches. Lisin *et al.* [7] show two methods where combining local and global features improved the accuracy of a classification task. Another aspect is the detection of *regional* features, where a region is defined as an arbitrary subset of the image. The extraction of Maximally Stable Extremal Regions (MSERs) by Matas *et al.* [8] highlights the advantage of region-based features: it produces both sparse and robust features that are particularly covariant to viewpoint change and affine transformations. Mikolajczyk *et al.* compare and evaluate a set of recent affine region detectors in [9], which is a main reference for the work proposed here.

In this paper, we present a novel region-based image feature detection approach using symmetric regions. Section 2 describes the regional symmetry feature detection. It is based on our earlier work on qualitative and quantitative symmetry measures proposed in [10] and [11]. After the detection of the features from the image data, we characterize each region by a SIFT descriptor [1] and subsequently match them using a common matching strategy. This is described in section 3. An evaluation and a comparison to other approaches is given in section 4, before we conclude our work in section 5.

II. SYMMETRY FEATURE DETECTION

A. Qualitative Symmetry

Motivated by psychophysical experiments on symmetry, we proposed a fast and compact one-dimensional operator to detect horizontal and vertical reflective symmetry features in earlier work [10]. Only pixels in the same image row or column are used for the detection of vertical or horizontal symmetry for a pixel p_i . It is not necessary to apply interpolation or trigonometric functions. The qualitative symmetry operator is based on the normalized mean square error function

$$s'(p_i) = 1 - \frac{1}{c \cdot m} \sum_{j=1}^m \sigma(j, m) \cdot g(p_{i-j}, p_{i+j})^2, \quad (1)$$

where m is the size of the operator applied to p_i . The normalization constant c depends both on the color space and on the radial weighting function $\sigma(j, m)$. The difference between two opposing points p_{i-j} and p_{i+j} is determined by a gradient function $g(p_{i-j}, p_{i+j})$.

This operator overcomes the problem of other symmetry detection methods that use large operators which are mostly unsuitable for robotic real-time tasks. Experiments show that our fast qualitative operator is able to robustly detect symmetry axis segments [12]. Like most other approaches, the operator shows the significant disadvantage of depending on an a-priori operator size m, i.e. a pixel's symmetry value is described with respect to a constant region around this pixel. Therefore, such *qualitative* operators return a relative, commonly normalized value of symmetry for each image element. This value describes the qualitative symmetry as low or high inside the constant region of size m. The lack of flexibility of those approaches is quite obvious. It would be more relevant to get quantitative information like the size of the symmetric region instead of its degree, with respect to a static operator size.

B. Quantitative Symmetry

For this purpose, a novel method to generate robust *quantitative* symmetry values was proposed in [11]. The approach is based on an algorithm computing bilateral quantitative symmetry information using an adopted Dynamic Programming technique referred to as Dynamic Programming Symmetry algorithm. For each image point, the pair of opposing image regions spans a single local search space. Each search space is computed to find an optimal mapping of the regions' elements. Symmetry information is finally extracted regarding the error of this mapping.

The optimal mapping and the overall error are computed in an iteratively growing subsquare of the search space. If the minimum error exceeds a given threshold in an iteration step, the calculation is aborted. The mapping end is returned using the search space indices $s_l(p_i)$ and $s_r(p_i)$ that now serve as a measure of symmetry. The environment given by $S(p_i) = s_l(p_i) + s_r(p_i)$ can be treated intuitively as the symmetric region around p_i . This operator offers quantitative, comparable symmetric range information for each image point. The disadvantage of this approach is the high effort in computing time, as a whole search space has to be treated for each pixel.

C. Symmetry Features

In this paper, we propose to combine qualitative and quantitative symmetry measurements into a mixed approach for fast and stable visual feature detection. To detect features from an image, we first use the qualitative operator to acquire fast symmetry information for each image point. Symmetry axis points are extracted by a line-independent maxima investigation on the symmetry data. This provides a binary representation of the axis points. The feature points are defined as intersections of vertical and horizontal symmetry axes. Therefore, we associate the axis points into straight line segments. This is done by a simple algorithm that searches for a segment start point and processes the line points to a segment end point. The segment representation gives access to additional information about each symmetry axis, e.g. length, orientation and maximum variance of the integrated points to the line segment. Segments with a large maximum variance correspond to curve segments. We iteratively split these at the point of maximum variance until they form straight subsegments.

Including quantitative symmetry measures now, each intersection of a horizontal and a vertical segment reveals an elliptical region feature $f_i = (\mathbf{y}_i, \theta_i, a_i, b_i)$ parametrized by

$$\mathbf{y_i} = (x_{\mathbf{y_i}}, y_{\mathbf{y_i}}),$$
 (center point) (2)

$$\theta_i = \frac{\theta + \theta}{2} - \frac{\pi}{4}, \quad \text{(orientation)} \quad (3)$$

$$a_i = \hat{S}_v(\mathbf{y_i}),$$
 (1st semi axis) (4)

$$b_i = S_h(\mathbf{y_i}),$$
 (2nd semi axis) (5)

where θ^h and θ^v correspond to the orientations of intersecting segments. Caused by line segmentation, intersections might miss the ideal symmetry maxima point, thus $\hat{S}_v(\mathbf{y_i})$ and $\hat{S}_h(\mathbf{y_i})$ are computed by finding the maximum vertical $S_v(\mathbf{x})$ and horizontal $S_h(\mathbf{x})$ in a small neighborhood of $\mathbf{y_i}$. Each feature ellipse can also be formulated as

$$F_{i} = \{(x, y) \in \mathbb{R}^{2} \mid A_{i}D_{i}(x - x_{\mathbf{y}_{i}})^{2} + 2B_{i}D_{i}(x - x_{\mathbf{y}_{i}}) \\ \cdot (y - y_{\mathbf{y}_{i}}) + C_{i}D_{i}(y - y_{\mathbf{y}_{i}})^{2} = 1\}, \quad (6)$$

where
$$A_i = a_i^2 \sin^2(\theta_i) + b_i^2 \cos^2(\theta_i),$$

 $B_i = (a_i^2 - b_i^2) \cdot \cos(\theta_i) \sin(\theta_i),$
 $C_i = a_i^2 \cos^2(\theta_i) + b_i^2 \sin^2(\theta_i),$
 $D_i = (a_i b_i)^{-2}.$

which is similar to the quadratic equation of central conics. Fig. 1(f) shows an example for the symmetry feature ellipses detected from the panoramic image sample in Fig. 1(a).

III. FEATURE DESCRIPTION AND MATCHING

In order to compare and match the perceived features in different views, each detected region has to be characterized by a significant descriptor. A number of feature descriptors have been elaborated in literature. A recent classification and overview of many techniques is given by Mikolajczyk and Schmid in [6]. For matching of symmetry features and for the comparison to a set of well-evaluated approaches from [9], all features f_i are described using a $4 \times 4 \times 8$ SIFT-descriptor [1]

$$d(f_i) = \{ d_u(f_i) \}_{u=1\dots 128}.$$
(7)

After the detection and description of symmetry-based regions, a measure of correspondence has to be defined to features that correlate the most. Since each feature is almost completely characterized by its descriptor vector, we therefore use the SIFT mapping

$$\rho(f_i, g_j) = \sum_{u=1}^{128} (d_u(f_i) - d_u(g_j))^2$$
(8)

to compute the similarity of two features f_i and g_j . The common application of feature matching is given by comparing a feature f_i from one scene with a set of features $\mathbf{g} = \{g_k\}_{k=1...n}$ detected in a second scene. The best mapping for f_i is thus given by

$$f_i \mapsto g_j: \quad g_j = \operatorname*{arg\,min}_{g_k \in \mathbf{g}} \rho\left(f_i, g_k\right).$$
 (9)

The most common type of matching that is applied by the SIFT operator [1] is the set of single best mappings

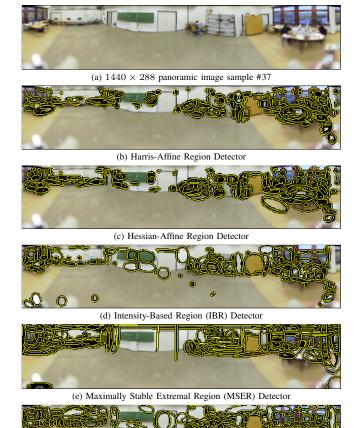
$$R(\mathbf{f}, \mathbf{g}) = \{\{f_i, g_j\} \mid f_i \mapsto g_j \land \rho(f_i, g_j) < t\},$$
(10)

where t describes a threshold, which we set to $\frac{1}{2}\rho(f_i, g'_j)$ with g'_j being the second best mapping for f_i . This ensures that the mapping is distinctive. The matching $R(\mathbf{f}, \mathbf{g})$ is not symmetric and usually non-injective, because more than one f_i may be mapped to one g_j . A common measure that is consulted to rate the quality of a matching between two images is the repeatability value. A basic the repeatability measure is defined as the ratio of the number of matches and the mean or minimum value of the number of features in both images. We use the minimum repeatability measure

$$r(\mathbf{f}, \mathbf{g}) = \frac{\|R(\mathbf{f}, \mathbf{g})\|}{\min(\|\mathbf{f}\|, \|\mathbf{g}\|)}.$$
(11)

IV. EVALUATION

In this section, we follow the experiments of affine region detectors in [9]. We evaluate the proposed symmetry feature detector in contrast to other well-elaborated feature detectors. For a set of panoramic images symmetry features are compared to Harris-Affine and Hessian-Affine Regions [6], Intensity-Based Regions (IBRs) [13] and Maximally Stable Extremal Regions (MSERs) [8]. While Hessian- and Harris-Affine are edge-based regions, IBRs, MSERs and the symmetry approach are oriented towards area-based regions. This can bee seen in the example images of Fig. 1.



(f) Symmetry Region Detector

Fig. 1. Sample features extracted by the five detectors: Harris-Affine and Hessian-Affine Regions [6], Intensity-Based Regions (IBRs) [13] and Maximally Stable Extremal Regions (MSERs) [8]. While Hessian-Affine and Harris-Affine are edge-based regions, IBRs, MSERs and Symmetry Regions are oriented towards area-based regions.

A. Region Size, Run-time and Feature Count

We calculate these features for the 1440×288 panorama image in Fig. 2 and present region size, run-time and feature count of the different detectors in Tab. I. The histogram of Fig. 3 shows distributions of image feature sizes, where the size of an elliptical region is computed as the mean value of its semi axes. Symmetry, MSER and IBR provide few and sparse features with mean feature size, while Harris-Affine and Hessian-Affine detect many small features.

For our symmetry detector, feature count and run-time do not depend on image size only, but also on symmetric image structure. The main effort is spent on the quantitative symmetry detection, where a growing search space for each image point is established. Larger symmetries cause larger search spaces and computation time. While symmetry features were computed on a Pentium M 1.7GHz Windows Laptop, standard Linux binaries¹ were used from [9] on a Linux PC* with a Pentium 4 3GHz for the other approaches. Though

¹www.robots.ox.ac.uk/~vgg/research/affine/detectors.html

MEAN SIZE, RUN-TIME AND FEATURE COUNT FOR ALL DETECTORS.detectormean sizerun-time#featuresSymmetry23.12~ 3.2 sec127

TABLE I

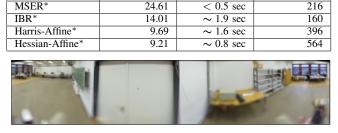


Fig. 2. 1440×288 panoramic image sample #1.

run-time of the detectors is not directly comparable because of the hardware differences, it can be concluded from the test that symmetry offers the most sparse set of features with large mean feature size. Additionally, the whole process of feature description and matching is depending on feature count, so symmetry features can be described and matched fastest.

B. Affine Transformation

Related approaches emphasize to be covariant under affine transformations like change of scale, rotation and perspective view. Covariance or overlap terms that elliptical representations of a feature cover the same image region in both images. The measure of quantitative symmetry intuitively illustrates the concept of scale robustness, as symmetry is highly proportional to scale. Only horizontal and vertical symmetry measures are used in our approach, thus the detection of features is not rotational invariant. Symmetry axes of horizontal and vertical operators are able to approximate slightly skewed axes of symmetry, but are rotational invariant for rotations of $\frac{r\pi}{2}$ only. In contrast to other approaches symmetry is comparatively weak in this covariance measure on affine transformations. Nevertheless, no multiple scale analysis or scale selection is needed, since scale emerges from symmetry. Symmetry features are not strongly covariant and fail in strict overlap measures, as it can be seen in Fig. 4.

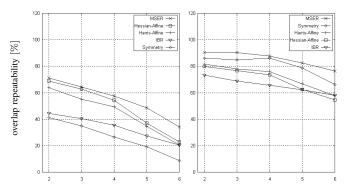


Fig. 4. Exemplary overlap graphs for the "wall sequence" from [9]. Left: Overlap of 40% error allowed. Right: Overlap of 80% error allowed.

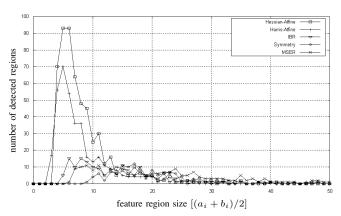


Fig. 3. Histogram of feature sizes for the panorama image #1 in Fig. 2.

C. Panoramic Warp

In contrast to the transformations (changes in image blur, scale, rotation, perspective view and JPEG compression) discussed in [9], panoramic warp is not an affine one. We exploit the properties of the detectors in an evaluation experiment on panoramic warp, which naturally includes changes in blur, scale, and panoramic view, as can be seen in the two panorama samples of Fig. 1(a) and Fig. 2. The two images are #1 and #37 of a sequence of 37 images that were recorded during a 3.38m straight movement using our robot platform. The images are retrieved from a SeiwaPro Panorama Eye[®] [14] omnidirectional vision system that is mounted above the stereo camera head of the robot. Like most catadioptric systems, the one applied here comprises a firewire color camera facing upwards to a hyperboloidal mirror surface. Omnidirectional images are restricted in image resolution, but offer the main advantage of providing a complete 360° visual perception of the surroundings in each time step. Methods for unwarping distorted omnidirectional views into panoramic views are well elaborated to offer user-friendly visual feedback. For the following experiments, we use image #1 as the reference frame for the other images #2 to #37.

1) Feature Matching: For each of the five detectors, $\mathbf{f}_{(1)}$ is computed, being the feature set for image #1. We also detect and describe the feature sets $\{\mathbf{g}_{(i)}\}_{i=2,...,37}$ to compute the matches between $\mathbf{f}_{(1)}$ and each $\mathbf{g}_{(i)}$. Hereby, we evaluate how sensitive the different detectors are with regard to different levels of panoramic image warp. The number of features and feature matches are shown in Fig. 5(a) and (b). We find that symmetry yields very few features and matches. To rate these matches, the repeatabilities $r(\mathbf{f}_{(1)}, \mathbf{g}_{(i)})$ (Eq. 11) are computed and plotted in Fig. 5(c). The plot presents clearly the repeatability decrease with larger distance because images differ more from the reference image #1 along the sequence. Additionally, it shows that the matching rates of MSERs and Symmetry Features are best to find correspondences from the detected features.

2) *Homographies:* However, detected matches are not always correct. There may be false positives, when image regions look the same. To distinguish between false and correct

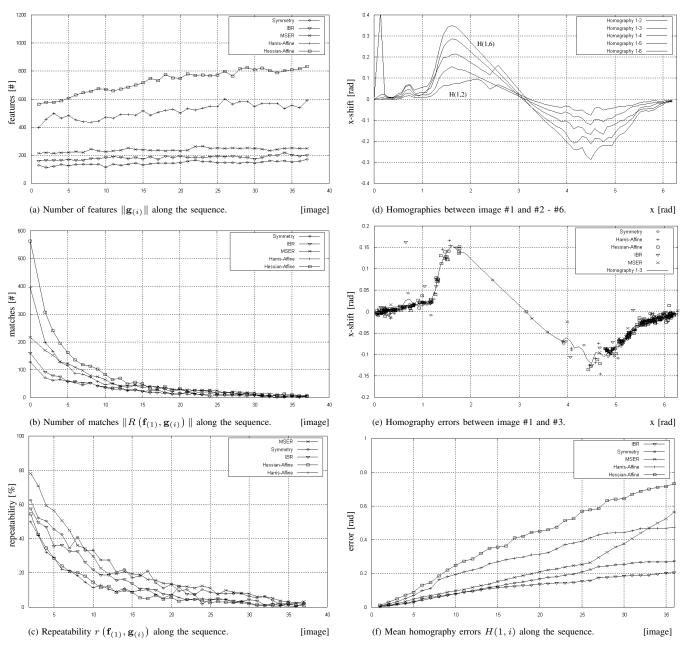


Fig. 5. For the comparison of all detectors, features of image #1 (Fig. 2) are matched to those of all other sequence images #2 to #37 (Fig. 1(a)).

matches, information about the exact image transformation is necessary. In [9], simple 3×3 homography matrices are used to define the ground truth of where a feature has to be after an affine transformation. On the one hand, panoramic image flow for robot applications is not an affine transformation. There are image regions that do not change (e.g. fixed robot parts in the image, regions along the axis of movement and regions that are far away) or others that warp in a nonlinear manner according to their size and their distance to the robot. On the other hand, the environment around the robot is unknown and dynamically changing, which makes panoramic homographies for robot applications impossible to establish. Therefore, we try to approximate each homography H(1, i) between image #1 and image #i by a column-based histogram of feature shifts. For each match that results from the feature matchings between $f_{(1)}$ and $g_{(i)}$, we assign its radial shift in x-direction to the column. If there are more shifts assigned to one column, the mean value is assigned. Note that all feature detectors' results are used to establish these homographies. Empty histogram cells are subsequently filled by interpolation. To handle outliers, each fifth entry of the histogram is used as a sampling point for a cubic spline that now describes H(1, i). Resulting homographies $\{H(1, i)\}_{i=2,...,6}$ are presented in Fig. 5(d). Regarding these homography graphs with increasing shift altitude and zero-crossings at the image edge (at 0 and 2π , respectively) and the image center (at π), one can obviously reason that the robot has moved away from a point in the image center. This is correct, as the robot moved a straight path from image #1 (Fig. 2) to image #37 (Fig. 1(a)).

3) Failure Analysis: After these two steps, we can compare the shifts of the feature matches to the corresponding homography for each image match. Fig. 5(e) presents the comparison between the $R(\mathbf{f}_{(1)}, \mathbf{g}_{(3)})$ of the different detectors and H(1,3). For the cause that homographies are acquired by the complete feature set, they are visibly influenced by these, but outliers are clearly recognizable. The largest outlier in the example in Fig. 5(e) can be detected at the left side of the image as a sample of the IBR method. Reviewing the image sequence, we find that this feature is one of the features describing a monitor screen. It has been matched to one of the other monitor screens in the image and truely is incorrect. In this homography H(1,3) there are few eye-catching outliers for IBR, Harris-Affine, MSER, Symmetry and Hessian-Affine.

The mean deviation of matches about the homographies along the whole sequence is depicted in Fig. 5(f). It can be seen that matching correctness decreases for each method the more the image #i differs from the reference image #1. Finally, we can conclude that IBR and Symmetry Features provide best matching correctness for the analyzed image sequence.

V. CONCLUSIONS AND PERSPECTIVES

In this paper, we proposed a new regional feature detection approach based on symmetry properties. We compared the approach to other state-of-the-art regional feature detectors for the task of panoramic image interpretation. The resulting symmetry-based feature regions are neither generally invariant nor covariant. Though the detector lacks in precision finding the same ellipse feature region after common image transformations, the applied SIFT descriptor is very robust to match corresponding features distinctively.

Run-time of the presented detector is comparatively high because of quantitative symmetry search space computation. While common qualitative symmetry detectors describe symmetry in a relative degree inside a window, the quantitative detector describes symmetry as a pixel range. Scale from quantitative symmetry is used to characterize the symmetric feature ellipse in extent only to speed up feature computation. Nevertheless, the detection of symmetric regions is sparse and well distributed in most real world samples, depending on image size and symmetric image content. Feature description and matching of symmetry features thus is faster than for other presented approaches that derive larger feature sets.

Each region covers a subset of the image that is significant related to quantitative symmetry, but not necessarily to object boundaries. The approach and the experiments presented in this paper show that scale from symmetry is successfully applicable as a modular inset for feature detection. No multiple scale analysis or scale selection is needed, as scale emerges from symmetry.

It was shown that the symmetry feature approach is well applicable for robust feature recognition during panoramic image warp. It offers comparatively few and significant features that support fast description and matching. Matched features are highly stable, distinctive and correct in combination with the SIFT descriptor. Another advantage of the symmetry approach is the strong relationship of features to the objects in the scene. Walls, doors, monitors and cabinets are frequently included by one feature.

An important step will be to increase the symmetry detector's rotation invariance and to thereby improve covariance and overlap. For this reason, it might be beneficial to find a symmetry orientation measure that yields a feature orientation before applying the twofold quantitative symmetry measures along and perpendicular to this direction. In this context, further evaluation of affine transformations is one topic of future work.

Other topics address the application of the developed symmetry features for visual robot navigation. The discussed homographies and shift values yet allow range estimation in combination with odometry. The object relationship will be a link to object classification to recognize doors or other common objects. Finally, a long-term goal is the autonomous robot localization and navigation in dynamic environments.

REFERENCES

- D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [2] K. Mikolajczyk and C. Schmid, "An Affine Invariant Interest Point Detector," in *European Conference on Computer Vision*, 2002, pp. 128– 142.
- [3] C. Harris and M. Stephens, "A Combined Corner and Edge Detector," in 4th ALVEY Vision Conference, 1988, pp. 147–151.
- [4] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of Interest Point Detectors," *International Journal of Computer Vision*, vol. 37, no. 2, pp. 151–172, 2000.
- [5] K. Mikolajczyk and C. Schmid, "Scale and Affine Invariant Interest Point Detectors," *International Journal of Computer Vision*, vol. 60, no. 1, pp. 63–86, 2004.
- [6] —, "A Performance Evaluation of Local Descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.
- [7] D. Lisin, M. A. Mattar, M. B. Blaschko, M. C. Benfield, and E. G. Learned-Miller, "Combining Local and Global Image Features for Object Class Recognition," *Proceedings of IEEE Workshop on Learning in Computer Vision and Pattern Recognition*, 2005.
- [8] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust Wide Baseline Stereo from Maximally Stable Extremal Regions," *Image and Vision Computing*, vol. 22, no. 10, pp. 761–767, 2004.
- [9] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, "A Comparison of Affine Region Detectors," *International Journal of Computer Vision*, 2005.
- [10] K. Huebner, "A 1-Dimensional Symmetry Operator for Image Feature Extraction in Robot Applications," *16th International Conference on Vision Interface*, pp. 286–291, 2003.
- [11] K. Huebner, D. Westhoff, and J. Zhang, "Optimized Quantitative Bilateral Symmetry Detection," *International Journal of Information Acquisition*, vol. 2, no. 3, pp. 241–249, 2005.
- [12] K. Huebner, "A Symmetry Operator and its Application to the RoboCup," in *RoboCup 2003: Robot Soccer World Cup VII*, ser. Lecture Notes in Artificial Intelligence, D. Polani, B. Browning, A. Bonarini, and K. Yoshida, Eds., vol. 3020. Springer, 2004, pp. 274–283.
- [13] T. Tuytelaars and L. J. Van Gool, "Matching Widely Separated Views Based on Affine Invariant Regions," *International Journal of Computer Vision*, vol. 59, no. 1, pp. 61–85, 2004.
- [14] SeiwaPro Co., LTD., "Panorama Eye R," www.accowle.com/english/index.html.