Local Structure Detection with Orientation-invariant Radial Configuration

Lech Szumilas[1], René Donner[1,2], Georg Langs[1,2], Allan Hanbury[1]
[1] Pattern Recognition and Image Processing Group
Vienna University of Technology, Favoritenstr. 9/1832, A-1040 Vienna, Austria
[2] Institute for Computer Graphics and Vision
Graz University of Technology, Inffeldgasse 16, 8010 Graz, Austria
{lech,donner,langs,hanbury}@prip.tuwien.ac.at

Abstract

Local image descriptors have proved themselves as useful tools for many computer vision tasks such as matching points between multiple images of a scene and object recognition. Current descriptors, such as SIFT, are designed to match image features with unique local neighborhoods. However, the interest point detectors used with SIFT often fail to select perceptible local structures in the image, and the SIFT descriptor does not directly encode the local neighborhood shape.

In this paper we propose a symmetry based interest point detector and radial local structure descriptor which consistently captures the majority of basic local image structures and provides a geometrical description of the structure boundaries. This approach concentrates on the extraction of shape properties in image patches, which are an intuitive way to represent local appearance for matching and classification. We explore the specificity and sensitivity of this local descriptor in the context of classification of natural patterns. The implications of the performance comparison with standard approaches like SIFT are discussed.

1. Introduction

Local image descriptors have proved themselves to be very useful for the recognition of objects in images. The "bag of key-points" [4] in combination with SIFT descriptors [7] are among the most successful techniques [5]. While the SIFT descriptor has been shown to perform well for region matching in transformed images [9], it has the disadvantage that it does not explicitly take the shape of the regions of interest (image patches) into account.

This paper presents a new local descriptor, the Orientation-invariant Radial Configuration (ORC) descriptor, which extracts shape properties of local image patches and their boundaries at the same time.

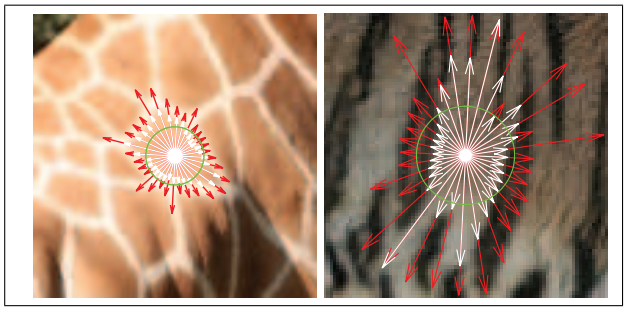

Figure 1. Example of boundary point detection: giraffe skin on the left and tiger skin on the right. The inner arrows (white) represent a geometrical description of the local structure interior, outer arrows (red) correspond to the local structure exterior.

One of the first attempts to capture the shape information in a local descriptor was proposed by Belongie and Malik [1], called shape context in the form of a log-polar histogram of the boundary points. Our detector is most closely related to the *Intensity-Based Method* of Tuytelaars and van Gool [12], which defines interest regions by detecting luminance transitions on rays emanating from local intensity extrema. Our approach differs in several ways. The main difference is that instead of fitting an ellipse to a detected region of irregular shape, the ORC descriptor encodes the shape. Furthermore, instead of detecting interest points at local intensity extrema, we use local symmetry extrema — the interest points then tend to appear in the center of salient image structures in the image. We also introduce an approach for detecting multiple pixel value transitions on the rays based on clustering. Finally, the proposed descriptor is able to encode multiple concentric structures in a single descriptor.

In this paper we focus on properties of the proposed ORC descriptor and compare ORC with SIFT performance for local structure matching in different images. In the following two sections we describe the nature of the symmetry based interest points and the construction of the ORC local de-

1-4244-1180-7/07/$25.00 ©2007 IEEE

scriptor. Distances between ORC descriptors are defined in Sec. 4.2. Sec. 5 presents the performance evaluation results.

2. Symmetry Based Interest Points

The final performance of any local descriptor is tightly associated with the positions of interest points. Since ORC is designed to capture the shape of basic local structures, which often correspond to patches of similar color pixels, we locate interest points in the centers of round/isotropic structures or along the symmetry axis of elongated shapes. Such positioning not only helps to detect local shapes but also makes descriptor matching more precise as will be shown in Section 4.2. For these reasons the detection of interest points utilized by the ORC descriptor is based on a radial symmetry measure and the interest points are aligned with local symmetry maxima. The radial symmetry measure is defined as:

$$S(x,y) = -\sum_{i=-R}^{R}\sum_{j=0}^{R} \|\boldsymbol{I}(x+i,y+j) - \boldsymbol{I}(x-i,y-j)\| \quad (1)$$

where $\boldsymbol{I}(x+i,y+j)$ is an image pixel intensity or color at coordinates $(x+i,y+j)$ and R defines the image window size used for symmetry measure calculation to be a $(2R+1)\times(2R+1)$ rectangle.

The symmetry measure S reaches a maximum (equal to 0) if all corresponding pixel pairs (x_c+i,y_c+j) and (x_c-i,y_c-j) are identical. It reaches a local maximum at the center of radially symmetric shapes (like a filled circle, star, etc.) or along the symmetry axis of elongated shapes. At the same time it reaches a local minimum along edges. Examples of interest points located at local maxima of the symmetry transform are shown in Figure 2.

The proposed symmetry measure tends to generate an excess of local maxima, but it does not miss any symmetrical regions in the image for the given scale of interest R. For our purposes, the quality of the interest points is sufficient over a wide range of R, which was set to $1/50$ of $\max(width, height)$ of the image.

Other symmetry transforms exist [8], intended primarily as salient feature detectors. They therefore do not detect as many interest points as are usually needed for representing the majority of basic local image structures.

3. Orientation-invariant Radial Configuration

The ORC descriptor produces a description of the local structure's boundary in the form of distances between the central interest point and N boundary points distributed at equal radial intervals of $2\pi/N$ radians around the interest point (see Figure 1). This approach allows the full description of simple convex-like shapes, both round and elongated. The description of complex structures is possible with multiple ORC descriptors.

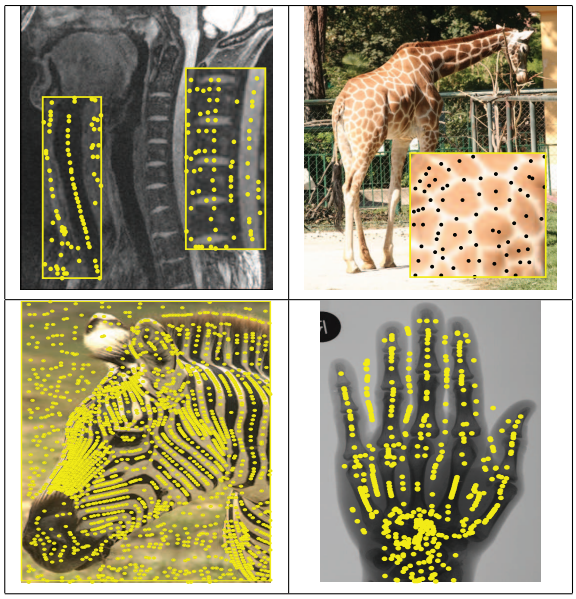

Figure 2. Examples of symmetry based interest points.

Descriptors are generated from circular regions surrounding the interest points, which must be large enough to enclose local structures. The initial size of the region r_{max} is estimated from the interest point adjacency. The circular region is divided into N equal sectors and luminance profiles are extracted along the radius of each sector. The method can be easily extended to use other features such as color information. The profile in each sector is then clustered into coherent and spatially contiguous regions while boundary points are associated with the region boundaries. This yields one or more boundary points in each sector (see Figure 3). Multiple boundary point configurations are then constructed of N points each, one from each sector.

3.1. Boundary Point Detection

Boundary points correspond to the edges or transitions between relatively different regions of pixels along the sector radius. The operation of boundary point detection is repeated for each sector separately.

The pixel grouping is based on an agglomerative hierarchical clustering of pixel related features (luminance), except that only spatially adjacent clusters can be joined into a node at the next clustering level. The mean feature value of cluster t at level l, containing the pixels at radial distances $r^l_{t,min} \le r \le r^l_{t,max}$ is denoted as C^l_t. At clustering level 1 each cluster contains exactly one pixel. The spatially adjacent clusters t and $t+1$ at level l are joined only if the following condition is fulfilled:

$$\|C^l_t - C^l_{t-1}\| > \|C^l_t - C^l_{t+1}\| < \|C^l_{t+1} - C^l_{t+2}\| \quad (2)$$

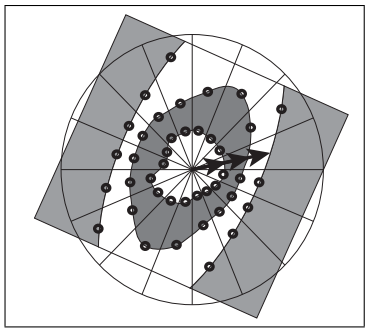

Figure 3. Multiple boundary points are detected along the rays. Arrows show three possible boundary point configurations corresponding to coherent image patches.

This way the clusters C_t^2 at clustering level 2 always represent adjacent pixel pairs or single pixels (if (2) was not fulfilled) and clusters at higher levels represent continuous sections of pixels along the radius:

$$C_t^l = \frac{1}{r_{t,max}^l - r_{t,min}^l + 1} \sum_{r=r_{t,min}^l}^{r_{t,max}^l} I_r \qquad (3)$$

where I_r is an image luminance value at radius r along the radial ray.

The operation of joining clusters is repeated until a complete clustering tree is built, containing two clusters at the top (see Fig. 4)

A boundary point should represent either an edge-like or a smooth transition between two regions with low pixel feature variance. In both cases we want that regions of relatively similar pixels are represented by a single cluster each. This implies that the position of the boundary points would correspond to the spatial boundary of the clusters. However, not every pair of clusters represents two relatively different regions if all clustering levels are considered. At lower clustering levels some cluster pairs may be a result of noise or weak transitions, while at higher levels even relatively strong transitions may disappear. For this reason only up to K cluster pairs exhibiting the strongest cluster transitions among all clustering levels are selected as locations of boundary points. The cluster transition $\Delta C_{s,t}^l$ between two clusters s and t at clustering level l is measured as an absolute difference between their intensity values:

$$\Delta C_{s,t}^l = \|C_s^l - C_t^l\| \qquad (4)$$

The cluster transitions are extracted only between clusters related to local extrema (see Figure 4), i.e. a minimum followed by a maximum pair or vice-versa. Such clusters do not have to be adjacent (see Figure 5), but no other extremum can appear between them.

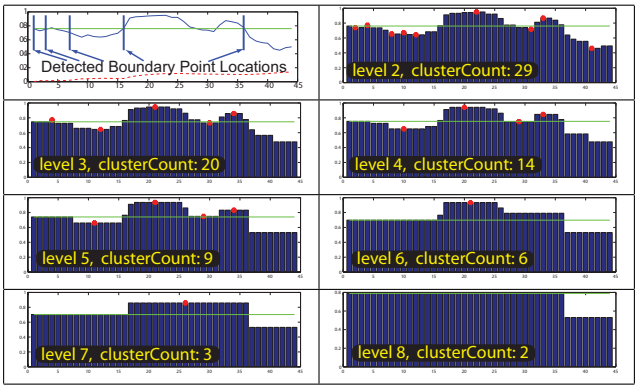

Figure 4. Example of pixel grouping along the radius in one of the sectors. The graph in the top-left corner represents the luminance profile along the sector radius. Other graphs represent grouping of the adjacent pixels at different clustering levels. Each bar represents one pixel, the bar heights correspond to the luminance.

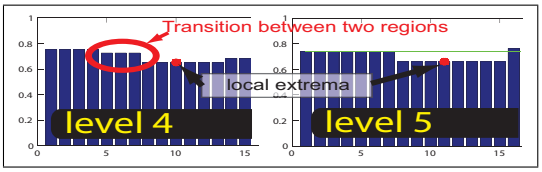

Figure 5. Example of the transition between relatively different regions.

The process of boundary point detection on each ray can be described by the following steps:

1. Perform hierarchical clustering of pixel feature values along the ray.
2. Find local cluster extrema at each clustering level.
3. Create a list of cluster transitions $\Delta C_{s,t}^l$ (using detected extrema) from all clustering levels.
4. Select up to K strongest cluster transitions and locate boundary points at the spatial boundary of related cluster pairs.

One of the advantages of the boundary point detection through hierarchical clustering is that each boundary point is associated with two clusters, which correspond to two regions separated by the luminance transition. The values of both clusters represent the mean luminance of these regions and are used for boundary point grouping in Section 3.2. Therefore each boundary point i is associated with two radii: q_i^1, which is the distance from the central interest point to the boundary of the local structure, and q_i^2, which covers also the second cluster ($q_i^2 > q_i^1$) (see Figure 1).

The boundary point detection can be compared to edge detection. The primary difference with resepct to typical edge detectors ([2] and [6]) is that we consider only a local

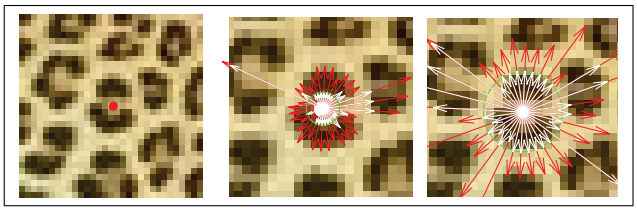

Figure 6. Example of multiple boundary point configurations for a patch on a leopard skin.

patch of the image and therefore estimate edge strength relatively to local conditions. It is possible to detect boundary points by simply using one of the typical edge detectors, but our experiments have shown that it is rather difficult to automatically adjust parameters to obtain results insensitive to noise but preserving perceptible edges across whole image. Our approach requires only a single parameter K, which in all experiments was set to 8 and produced consistent results.

3.2. Boundary Point Grouping

The appearance of natural objects frequently exhibits high complexity. To take advantage of this information a grouping of boundary points between adjacent rays is necessary, whereby a compact and continuous representation is achieved. Figure 3 shows an example of multiple local structures present around a single interest point. It is possible to detect them by grouping boundary points according to their similarity. Their correct detection depends on the choice of a similarity measure. For natural scenes with uniform color patches, color distance gives good results.

A more robust strategy is proposed, based on the grouping of those boundary points in adjacent rays, exhibiting similar boundary luminance transition and similar inner luminance spread i.e. luminance standard deviation along the radius between interest point and boundary point. The boundary luminance transition is the cluster transition (Equation 4) associated with the boundary point. The inner luminance spread helps to capture multiple structures (see Figure 6) or when weak repetitive patterns/textures are present between interest point and boundary points. Both values, which constitute the *boundary point features*, are normalized by the average of luminance transitions and spreads of all boundary points.

The process of boundary point configuration detection can be summarized as follows:

1. Extract features for all boundary points in N sectors.
2. Perform hierarchical clustering of boundary point features in adjacent sectors (see Figure 7).
    - A single cluster cannot contain multiple boundary points from the same sector. This implies that the maximum number of boundary points within any cluster cannot be higher than the number of sectors.
    - The clusters cannot contain more than N boundary points, which is equivalent of a closed curve.
3. Extract all clusters B_k containing at least N_{min} boundary points. All other clusters are rejected.
    - The parameter N_{min} is typically set to $0.25N$ in order to discard small structures, which are usually less distinctive.

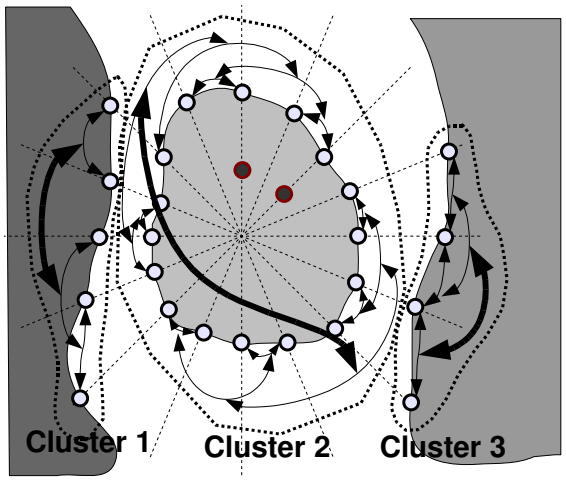

Figure 7. Example of boundary point grouping. Thin arrows show grouping of boundary points and clusters in adjacent sectors. Thick arrows show the top level clusters of the clustering tree (no more cluster merging possible).

The boundary point grouping may yield several boundary point configurations, each separately used during the matching of two descriptors (see Figure 6).

The performance of this strategy was verified experimentally on a database of natural scenes as well as medical images of MRI scans and gave consistent results (e.g. Figure 6). Other grouping strategies, including the use of distance between central interest point and boundary points as an additional feature as well as local sector grouping will be investigated in the future.

3.3. Refining Scales in Low Gradient Sectors

The ORC descriptor is extracted from the finite circular image region. The size of the region is derived from the interest point distribution in the neighborhood and encloses basic local structures in almost all cases. However, it is not always possible to automatically find a sufficient region size for extremely elongated structures. The boundary of an elongated structure in the structure dominant orientation may fall outside the circular region, which makes estimation of boundary points problematic and leads to random

results i.e. they are associated with the weak edges or transitions and not with the real structure boundary. However, the luminance spread ν_i^k calculated over the range of radii $0..q_i^2$ (covers structure interior and exterior) is significantly lower for "weak" boundary points than for the points associated with the real structure boundary. This measure can be used to both detect and refine "weak" boundary points. The i-th boundary point, belonging to the k-th configuration is classified as a weak one if:

$$\nu_i^k < 0.5\overline{\nu_{j\in B_k}^k}. \tag{5}$$

where the term $\overline{\nu_{j\in B_k}^k}$ corresponds to the luminance spread average over all boundary points within the k-th configuration. The consistent detection of "weak" boundary points using condition (5) is possible because the majority of boundary points, even in elongated shapes, usually correspond to real structure boundary, as in the example in Figure 1. The robustness of this condition was verified experimentally.

The "weak" boundary points are simply removed, if other boundary points exist in the corresponding sector, or otherwise q_i^1 of the point is increased until condition (5) is met or r_{max} is reached.

4. Descriptor Distance

We discuss the the measurement of scale and rotation invariant distance between ORC descriptors.

4.1. Scale Invariant Descriptor Distance

ORC distance is calculated between two boundary point configurations $f(i)$ and $f(j)$. Each configuration is represented by an M row by N column matrix, where N is the number of sectors and M is the number of features per sector. In the subsequent evaluation N was set to 32, representing a sensible trade off between computational complexity and the level of details captured by the descriptor. We have used two features ($M = 2$) per sector, representing the boundary point radii values q_i^1 and q_i^2.

The Euclidean distance between descriptor i and descriptor j, rotated by p sectors is calculated accordingly to:

$$d(i,j,p) = \sum_{n=1}^{N}\sum_{m=1}^{M}\left(f(i)_{m,n} - \varsigma_{i,j}(p) f(j)_{m,(n+p) \bmod N}\right)^2 \tag{6}$$

where $\varsigma_{i,j}(p)$ is a relative scale between descriptors i and j, which is discussed later in this section. Equation (6) is equivalent to the horizontal rolling of the descriptor j matrix p times to the right.

Local image descriptors, such as SIFT, are typically extracted from an interest region, which provides an estimation of the local image structure scale. While this approach allows the simplification of descriptor extraction and matching complexity it also makes a descriptor more sensitive to inaccuracy of the scale or interest region detector.

In our approach we have decided to avoid prior scale detection and replace it with the detection of relative scale between two descriptors, which is calculated during descriptor matching. The relative scale $\varsigma_{i,j}$ between descriptor i and j, for a given rotation p of the descriptor j is calculated as follows:

$$\varsigma_{i,j}(p) = \frac{\sum_{n=1}^{N} f(i)_{1,n} f(j)_{1,(n+p) \bmod N}}{\sum_{n=1}^{N} \left(f(i)_{1,(n+p) \bmod N}\right)^2} \tag{7}$$

It can be easily proven, that the distance (6) reaches minimum, for a given rotation p, if descriptor j is scaled by $\varsigma_{i,j}(p)$.

4.2. Rotationally Invariant Distance

Rotational invariance can be achieved by estimating local dominant orientation, either from local gradient covariance or by fitting an ellipse to the ORC boundary. This strategy works well for elongated or anisotropic structures but may produce unstable results for structures with no dominant orientation. We therefore apply a rotationally invariant distance measure, which works well in both cases:

$$d(i,j) = \min_p(d(i,j,p) : p = 1..N, M = 1) \tag{8}$$

As in the case of scale invariance we opt rather for detection of relative orientation between two descriptors than prior orientation estimation for each descriptor. The relative orientation between two configurations p_{min} is found from the minimum configuration distance calculated for $M = 1$ (taking into account only q_i^1 values).

The distance between two boundary point configurations depends to a large extent on the consistency of interest point positioning. The distance calculated between two ORC descriptors originating from the same structure but calculated at different interest points is not equal to 0. This is one of the reasons that the ORC descriptors are calculated around symmetry based interest points, which produce highly consistent results. It is possible however to avoid this dependency of the ORC distance calculation on the initial interest point location, by performing an iterative optimization of the distance, where one of the descriptors is gradually translated to minimize the final distance. While our experiments show this solution works well it also significantly increases computational complexity. The results of ORC testing presented in Section 5 were generated without refining interest point positions.

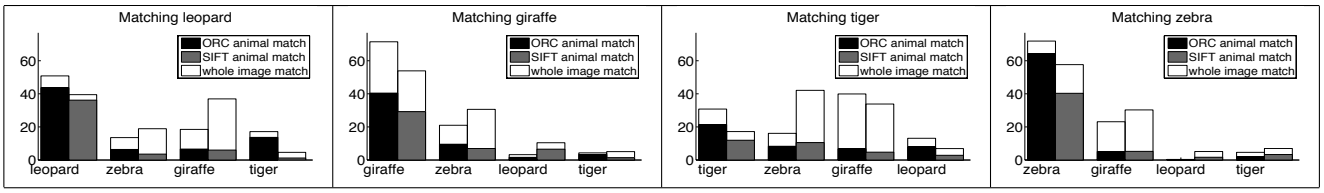

Figure 8. Animal matching results per animal category. Black and gray bars show the percentage of category descriptors matched inside corresponding animals for ORC and SIFT respectively. White bars show the percentage of category descriptors matched anywhere in corresponding animal images..

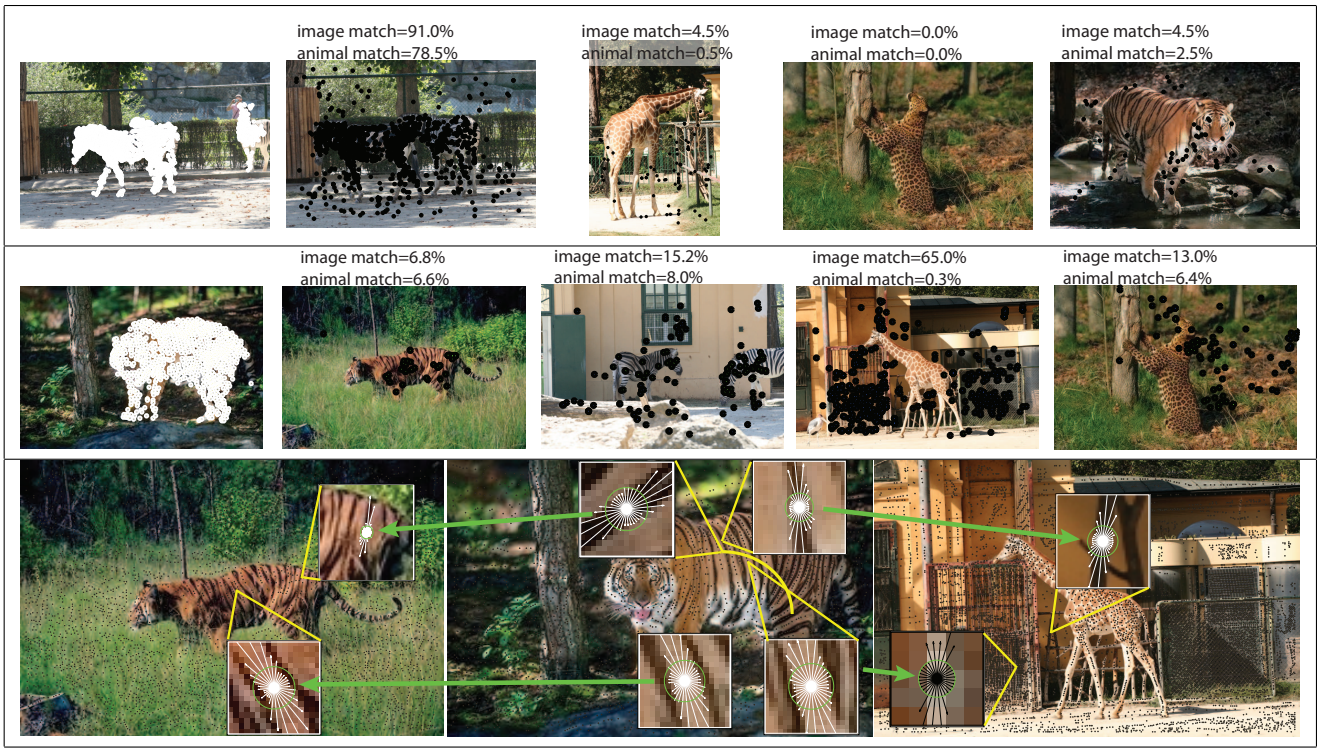

Figure 9. Example of descriptor matching in the image sequences. Black dots show closest matches to the descriptors originating from the zebra (top row) and the tiger (middle row) in the first image to the left (white dots). The bottom row shows examples of ORC descriptors originating from the tiger (in the middle) matched to the descriptors in the other two images. Some of the local structures in the right image are similar to the tiger stripes on a local scale.

5. Local Structure Matching - Performance Evaluation

The proposed testing procedure evaluates the matching of local structures found in the following animals: tiger, zebra, giraffe and leopard. Tigers and zebras contain mostly elongated features e.g. straight and bent stripes, while leopards and giraffe contain a mixture of round and elongated shapes. Every animal type represents a distinctive pattern of relatively low complexity local structures, suitable for matching using local descriptors.

Category	Group count	ORC (%)	SIFT (%)
leopard	10	43.6	36.2
giraffe	9	40.4	29.2
tiger	9	21.4	12.0
zebra	8	64.1	40.3

Table 1. Animal matching results per animal category.

Image descriptors are often evaluated by how well they match structures between an image of a scene and a transformed image of the same scene [9]. For the evaluation presented here, we wish to test how well a descriptor can generalize. That is, it should detect the same type of structure in different scenes, e.g. the giraffe skin pattern in many images of different giraffe.

The test procedure operates on groups of images, each containing five images with different animals and backgrounds. The images contain only one type of animal each

and were manually segmented to separate the background from the animal. The first two images in the group contain an animal of the same type and the remaining three images contain other animals e.g. (zebra 1, zebra 2, giraffe, leopard, tiger). Descriptors are extracted for the region corresponding to the animal (obtained from the manual segmentation) in the first image in the group. These descriptors are then matched with all descriptors from the rest of the images in the group to find the closest match for each descriptor. This way the percentage of closest matches inside the same type of animal as well as inside other animal types and backgrounds can be recorded.

The results of local structure matching were obtained from 36 image groups (180 images). The image groups are divided into 4 categories, each containing one type of animal occurring twice in every group. Figure 8 shows the average percentage (within each category) of descriptors from the animal in the first image matched inside animals in other images in the group (black and gray bars) as well as matched anywhere in other images (white bars). A more detailed version of these results is available in [11]. The summarized results, showing only the percentage of descriptors matched in the same kind of animal are presented in Table 1.

The results expose limitations of SIFT and ORC, primarily associated with their locality and the fact that both can discriminate between relatively simple image structures. Images used for testing (see [11]) contain a variety of backgrounds and as one could expect there is high chance of finding local structures in one of the image backgrounds that are similar to the structures present in the animal e.g. most of the tiger descriptors are matched with the fence in the background of the image containing a giraffe in Figure 9.

We compare the results of the proposed ORC descriptors with SIFT descriptors, where the SIFT descriptors and interest points were generated using code available from [13].

6. Conclusions

The evaluation of ORC and SIFT descriptors applied to the local image structure matching has shown that both descriptors are comparable with ORC performing 10-20% better than the standard SIFT in the experiments on images of four types of animal. It has also shown that the applicability of both local descriptors to a general object recognition tasks is limited by the lack of global context. Significant improvements are possible only if the co-occurence and spatial relationship of local structures are taken into account.

Various attempts have been made to add global context to the SIFT descriptor [10] or introduce spatial configurations of local descriptors [3] in order to better capture larger image structures or disambiguate multiple regions with locally similar appearance. The ORC descriptor provides a natural way for constructing spatial configurations of multiple descriptors, since adjacency of descriptors can be easily extracted from the boundaries shared by neighboring structures. This is a topic of future research.

7. Acknowledgements

This work was partly supported by the European Union Network of Excellence MUSCLE (FP6-507752), and the Austrian Science Foundation (FWF) under grant SESAME (P17189-N04) and AAMIR (P17083-N04). We would like to thank Philipp Peloschek and Klaus Friedrich from the Vienna Medical University for providing medical data.

References

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

[2] J. Canny. A computational approach to edge detection. *IEEE Trans. PAMI*, 8(6), 1986.

[3] G. Carneiro and D. Lowe. Sparse flexible models of local features. In *in Proceedings of ECCV'06*, 2006.

[4] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray. Visual categorization with bags of keypoints. In *Workshop on Statistical Learning in Computer Vision (at ECCV)*, 2004.

[5] M. Everingham, A. Zisserman, C. Williams, and L. V. Gool. The Pascal visual object classes challenge 2006 (VOC2006) results. Technical report, PASCAL Network of Excellence, 2006.

[6] M. Heath, S. Sarkar, T. Sanocki, and K. Bowyer. A robust visual method for assessing the relative performance of edge-detection algorithms. *IEEE Trans. PAMI*, 19(12):1338–1359, 1997.

[7] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

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

[9] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Trans. PAMI*, 27(10):1615–1630, 2005.

[10] E. N. Mortensen, H. Deng, and L. Shapiro. A SIFT descriptor with global context. In *Proceedings of the CVPR*, volume 1, pages 184–190, 2005.

[11] L. Szumilas. Supplemental test data. http://www.prip.tuwien.ac.at/ lech/cvpr2007data.pdf.

[12] T. Tuytelaars and L. van Gool. Matching widely separated views based on affine invariant regions. *IJCV*, 59(1):61–85, 2003.

[13] A. Vedaldi. An open implementation of SIFT. http://vision.ucla.edu/ vedaldi/code/sift/sift.html.