

Shape Centered Interest Points for Feature Grouping

David Engel

david.engel@tuebingen.mpg.de

Cristóbal Curio

cristobal.curio@tuebingen.mpg.de

Max Planck Institute for Biological Cybernetics

Spemannstr. 38

72076 Tübingen, Germany

Abstract

Image encoding using interest points is a common technique in computer vision. In this paper we present a scale and rotation invariant shape centered interest point (SCIP) detector. By means of detecting singularities in Gradient Vector Flow (GVF) fields we find points of high symmetry in the image. Due to the nature of the underlying GVF field we can employ our features to group together edge-based interest points such as SIFTs. This feature grouping provides a strong descriptor for SCIPs and can help to encode valuable information about the image for computer vision tasks. We demonstrate the main properties of our features such as scale and rotation invariance and further robustness against noise and clutter in a series of experiments. We show that the information they encode is to a certain degree complementary to SIFT. Furthermore, we evaluate them in an edge map reconstruction task to assess the amount of image information they encode. Finally, we show the power of feature grouping with our framework in a multi-category classification task on natural images from the StreetScenes database.

1. Introduction

Image encoding is a vital step for any computer vision algorithm. Since pixels are only an artifact of image quantization they are by themselves generally not very informative. As a consequence, encoding an image purely by its color values is not useful for most tasks, thus calling for more compact image representations. It is common practice in computer vision to represent the information contained in an image using interest points. Interest points are formed at locations where the local image structure has special properties. For most computer vision tasks it is important to have interest points that are invariant against rotation and scaling and robust against noise and background clutter. Running computer vision algorithms on a set of interest points

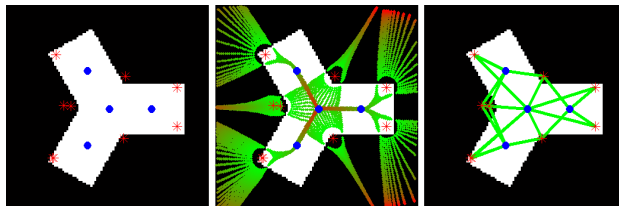


Figure 1. **Left:** Example with SCIP (blue) and SIFT (red) interest points. **Middle:** Evolution of the traces $T_{i,n}$ overlaid on the image (Color gradient green to red indicates iteration steps of the traces). **Right:** Grouping of the SIFTs to the SCIPs (green lines).

derived from an image is a widespread method that yields substantial speedups. The rationale behind this approach is that, since well chosen interest points are located at all informative image locations for the task at hand, the output of the algorithms will be the same whether it is computed densely on the whole image or only sparsely at interest points. Depending on their nature interest points can be located along the outline of shapes or they can be located inside the shape or object of interest. Edge-based interest points are often called corner points and have proved to be very useful for many computer vision tasks. For segmented images, corner interest points offer a very powerful way to encode an image. Natural images that contain background clutter on the other hand pose a harder problem. Corner interest points are located on or close to the border between two shapes, thus it is likely that they contain information about the shape of interest as well as about background clutter. Differentiating between those two is a hard and often ambiguous task.

A class of interest points that is less prone to this problem are shape centered interest points, often also called medial features. They occur at image locations where the local image structure is highly symmetrical. As they are located inside a shape they are less likely to be disturbed if the background outside of the target shape changes. Several ideas to identify such medial features have been proposed recently.

In the presence of noise and clutter the information available at individual interest points can be seriously degraded.

To address this problem we perform a bottom-up grouping in which interest points that are likely to belong to the same shape are clustered together (see Figure 1). Such grouped features contain far more information about the local shape and provide robustness against noise and clutter. Our image feature grouping takes advantage of a Shape Centered Interest Point detection scheme, that we want to call SCIP from here on. SCIP is a symmetry detection scheme that is based on PDEs, *i.e.* the Gradient Vector Flow (GVF) field [1]. Our algorithm is based on the medial feature extraction pipeline by Engel and Curio [2], which we have adopted here for the implementation of this novel feature grouping approach. After discussing related work in Section 2 we provide details of our implementation in Section 3. In Section 4 we demonstrate with three experiments the advantage of our approach by 1) showing the stability of SCIPs in location and scale estimation under noisy image conditions, 2) demonstrating the complementary information they encode in comparison to SIFT [3] interest points for an edge reconstruction task and a patch based similarity computation, and finally, 3) demonstrate the usefulness of employing SCIP for grouping edge features in a classification tasks.

2. Related Work

Features of intermediate complexity have been recognized as being useful entities for both human object recognition and in computer vision systems [4][5][6][7].

In biologically plausible computer vision systems features of intermediate complexity have been learned from low-level image features [8], building upon the biologically motivated approach of Riesenhuber and Poggio [9] for hierarchical object recognition. Gestalt law principles have been explicitly implemented for grouping features within recognition architectures, *e.g.* [10]. A similar approach has been suggested that incorporates these principles into the design of Markov Random Field structures [11] that are based on the FORMS shape recognition system [12]. There exist a wide variety of approaches to feature and edge grouping, *e.g.* the contour completion work of [13] and [14].

Ridge like basis-function as intermediate generative features have emerged from principles of unsupervised learning, *e.g.* [15][16]. As a more general concept, the strength of the medial features in vision has been emphasized by Kimia and others, *e.g.* [17][18]. In this line of research wave propagation models based on the solution of the Eikonal equation in the image plane have been employed to derive features for shape matching tasks. Alternatively, GVF [1] fields have been recognized for the computation of medial features in cluttered scenes [2]. The GVF approach made many vision tasks amenable for cluttered environment. Initially developed for long range contour alignment, GVF has recently been investigated for shape char-

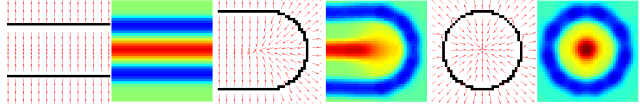


Figure 2. Three classes of medial features: Ridge, End-Stop and Circular. Shown are the normalized GVF fields V_N and the resulting flux flow \mathcal{F} (see Equation 1)

acterization of segmented silhouettes [19][20]. These approaches extend the notion of medial features beyond blob or ridge-like structures [21] as they offer more descriptive power. Lu *et al* proposed the M-Rep model [22] that has been successfully applied to even 3D shape analysis and registration. This approach particularly learns the statistics of so called medial atoms which are used for similarity judgments of graph model structures. Scale-invariant coding of image features has been addressed since the invention of, *e.g.*, SIFT features and region based detectors (*c.f.* [23]). Latter has been incorporated for scale invariant object recognition [24]. In general, the development of region-based image coding schemes received increased attention as they proved to be very useful to model the interplay of segmentation and recognition [25][26][27].

3. Methods

In this Section we describe how the shape centered interest points are generated and how we estimate a robust local scale and orientation for each interest point. Furthermore, Section 3.3 details how we can exploit the normalized GVF field to group edge-based interest points at our medial features. Figure 3 illustrates the full feature extraction and grouping pipeline.

3.1. Constructing Shape Centered Interest Points

By finding locations of high symmetry in the image we obtain shape centered interest points (SCIP). Such points are only influenced by the enclosing shape and are therefore invariant against background edges on the outside providing them with stability against clutter. A common technique to detect symmetries in the image is by looking for local maxima of the distance transform D_T which assigns the distance to the nearest edge in pixels to each point in the image. A major drawback of using D_T is its instability in the presence of noise (*c.f.* the evaluation in Section 4.1). To tackle this problem we employ the GVF field by Xu and Prince [1] which has been shown to provide a high stability against noise.

GVF has been developed as an image force for long range contour deformations. The goal is to find a smooth vector field $V(p) = [u(p), v(p)]^T$ that is aligned with the gradients of edge map ∇f where $p = (x, y)$ is a point in

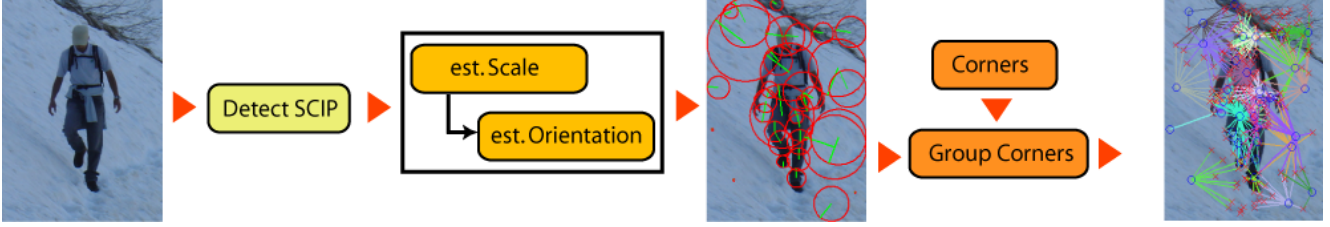


Figure 3. Feature extraction and grouping pipeline. For an image, SCIPs are detected, and a local scale and orientation estimation is performed to obtain rotation and scale invariant feature points. In the next step the linking procedure is applied to group together corner interest points at the SCIPs. Different colors (right) correspond to different groupings. The sample image has been downscaled to reduce the number of features for visualization purposes.

the image \mathcal{I} . To this end $V(p)$ is iteratively optimized to minimize the functional \mathcal{E} using variational calculus

$$\mathcal{E} = \int \int \underbrace{g(|\nabla f|) |V - \nabla f|^2}_{\text{data term}} + \underbrace{h(|\nabla f|) \nabla^2 V}_{\text{smoothing term}} dx dy.$$

The data term penalizes vector fields V that differ from the gradients in the edge map. The smoothing term acts as a regularizer that suppresses sudden changes in the GVF field. To determine the trade-off between these two conflicting goals the two gating functions g and h have been introduced. They are designed to be complementary, enforcing strict adherence of V to the image gradients near high edge magnitudes while extending the gradient map smoothly farther away from the edges. We normalize the solution $V(p)$ at each point in the image and obtain $V_N(p) = V(p)/\|V(p)\|$. Using $V_N(p)$ allows us to perform all consequent operations invariant with respect to the local magnitude of the vector field. For examples of the normalized GVF field see Figure 2.

We interpret $V_N(p)$ as the gradient of the L_2 -norm of a noise free distance function $V_N(p) \approx \nabla D_T(p) \forall p$. Based on this interpretation we observe that the ridges of distance function $D_T(p)$ coincide with singularities or shock points in the flow field $V_N(p)$. Following the idea of Pizer et al [28] we detect singularities of $V_N(p)$ by calculating the divergence of $V_N(p)$ approximated by a ring integral. We obtain the flux flow \mathcal{F} for each point p in the image

$$\mathcal{F}(V_N(p)) = \text{div } V_N = \frac{\oint \langle V_N, \mathcal{N} \rangle ds}{\text{Area}}, \quad (1)$$

where \mathcal{N} denotes the normals on a ring around p with diameter of 7 pixels through which the flux flow \mathcal{F} is computed. The computation of this ring integral can be implemented efficiently as a convolution of the GVF field with a precomputed kernel containing the normal vectors of the ring. High values of \mathcal{F} indicate locations of high symmetry while high negative values indicate the location of salient edges in the original image.

Using the flux flow field \mathcal{F} we sample SCIP locations using a nonmaximal suppression scheme which selects local maxima of \mathcal{F} . To avoid too dense sampling along ridges we enforce a distance constraint that ensures that interest points are at least 3 pixels apart from each other. We keep only interest points above a flux flow value θ which controls the number and kind of interest points.

We can roughly categorize three main classes of SCIP interest points that can be identified by their flux flow values as shown in Figure 2. In our implementation the flux flow ring integral yields the highest flux value for circular shapes. At semi-circular locations of the edge map so called end-stop points are formed which lead to distinctly lower flux flow values. A third class of interest point with even lower flux flow values is formed along ridges. Transitions between these three classes like tripods or cone-like structures have intermediate values.

3.2. Scale and Orientation Estimation

SCIP features are tightly connected with the surrounding shape and thus offer the possibility to estimate a meaningful local scale as has been shown by us in [2]. The features are centered between at least two salient edges. Locating these edges relative to the interest point provides a stable and interpretable local scale estimate. Corner based interest points also offer ways to estimate a local scale. However, as corners are by definition located along the edges between at least two regions the estimated scale is not associated with a particular shape and consequently not directly interpretable.

The most direct way to find the scale at an interest point is by calculating the mean edge energy on a ring centered at the interest point. This mean energy, computed as a function of the ring radius, is expected to have a maximum as it crosses the salient edges that were involved in the formation of the interest point. This method can become unstable in the presence of noise and clutter. To overcome this we extend the approach of [2]. Using the normalized GVF field V_N we calculate the scalar product between vector field V_N and a disc with circularly inward pointing normals for different disc radii. The estimated local scale is the first maxi-

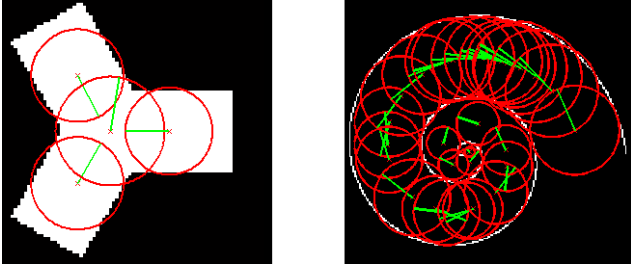


Figure 4. Examples of medial feature interest points with estimated local scales and orientations computed at one image scale.

mum of the function $\mathcal{S}(r)$

$$\mathcal{S}(r) = \frac{\int_0^{2\pi} \int_0^r \langle V_N(\alpha, r), \mathcal{N}(\alpha, r) \rangle dr d\alpha}{\pi r^2}, \quad (2)$$

where r is the radius of the disc of normals $\mathcal{N}(\alpha, r)$. Inside a shape the flow fields point toward the interest point leading to positive contributions to \mathcal{S} . By computing the interest points along a pyramid of scales we can detect them across a wide range of image resolutions in a scale space and thus gain further scale invariance.

Another desirable property of interest points is rotation invariance. We obtain it by estimating an orientation at the interest points. We compute the first eigenvector of the autocorrelation matrix of the thresholded flux flow field \mathcal{F} . Examples of the output of the feature extraction pipeline including the estimated scales and orientations are illustrated in Figure 4. At rotation symmetric SCIPs (*e.g.* circular shapes, or ridges) the estimated rotation may not be unique. But, since all possible orientations obtained this way have the identical flux flow signatures they can all be recognized equally well and all orientations have to be assumed to be correct.

3.3. Grouping edge-based features

We now describe how the shape centered interest points can be used to group together corner based interest points such as SIFT. We can use the underlying GVF field to link the corner-based SIFT interest points with their strong edge-descriptors to the medial features. We show that this linking yields an even stronger descriptor for the shape centered interest points in Section 4.3. We compare the power of the linked descriptors to the information contained at the corner points alone in a multi-class classification task.

Corner interest points are formed at locations with a rich edge structure thus providing very powerful and uniquely identifiable descriptors. However, they suffer from the drawback that they are located at the edges between two shapes making it hard to associate them with one shape in, *e.g.*, a classification task. Our shape-centered interest points have the benefit of being tightly associated with one specific

shape but suffer from the drawback that there is only scarce local edge information. By combining the advantages of both approaches we can provide strong descriptors at locations that belong uniquely to one shape.

The normalized GVF field V_N provides a unique gradient direction for each image location (see Figure 2). The shape-centered interest points are located at the shock-loci of the field. Starting at corner interest points we follow the vectors in the GVF field (*c.f.* Figure 2). This procedure will converge to the local minima of the flux flow field \mathcal{F} .

Equation 3 demonstrates the idea of iteratively following the normalized GVF field using a trace T (*c.f.* Figure 1). Here, a trace is a series of $2D$ image positions from a corner to a SCIP. Because there might be multiple medial feature interest points associated with a corner we have to follow the GVF field in all directions. To reach all adjacent minima of \mathcal{F} we generate 36 traces $T_{i,0}$ ($i \in 1..36$) at 10° intervals on a ring with a radius of 5 pixels around the corner point:

$$T'_{i,n+1} = T_{i,n} + V_N(T_{i,n}), \quad (3)$$

where $T'_{i,n+1}$ is the updated position of the i th trace after $n+1$ iterations. Because of numerical instabilities or noise the traces might not converge to the SCIPs. To address this we connect the corner interest point to the medial feature as soon as one of the traces arrives in the disc with the radius of the estimated local scale \mathcal{S} around the interest point. Figure 1 shows traces of SIFT interest points in the GVF field while moving toward the medial features.

This grouping algorithm usually converges after 30 – 50 iterations. The scheme can be significantly sped up by terminating traces that have reached a local minimum and by combining traces that have merged (*i.e.* delete $T_{i,n}$ if $T_{i,n} = T_{j,k}$ for any $k < n$ and $i \neq j$). With this algorithm we obtain a bipartite graph of groupings between corner and medial feature interest points.

4. Evaluation

In this Section we present three experiments highlighting the properties of our features and their use as a tool for feature grouping. First, we show the robustness of the SCIP features against noise and clutter. To show that they can encode relevant information which is qualitatively different and complementary to the information encoded by corner based interest points we evaluate both kinds of features in an edge map reconstruction task. Finally, we demonstrate the power of our feature grouping approach in a classification task on natural images.

4.1. Experiment 1: Robustness against noise and clutter

First, we want to demonstrate that our SCIP features possess a higher robustness against noise than the standard

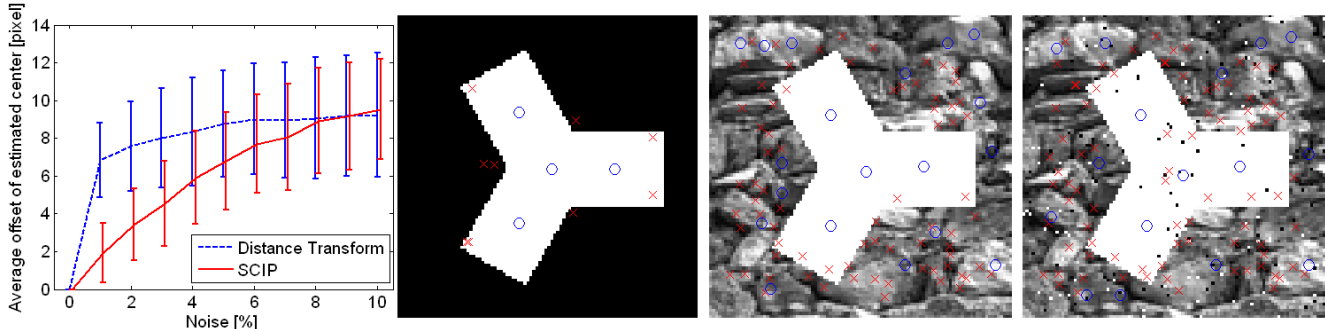


Figure 5. **Left:** Comparison of robustness between the normalized GVF field and normal distance transform against noise. Percentage of noise pixels along the x-axis. The y-axis shows the average distance and standard deviation of the estimated center of symmetry to the real center of a circle. **Right:** Robustness of the medial features against clutter. Blue circles denote shape centered interest points while red crosses are SIFT points. The three images show: No clutter, Clutter, and Clutter plus salt and pepper noise.

method of finding symmetry points by detecting the local maxima of the distance transform D_T . As task we chose the detection of the symmetry point in a circle with radius 25 pixel (*c.f.* Figure 2 right). We start with an edge map of the circle and add different amounts of noise by randomly setting pixels in the edge map to 1. The maximum of the flux flow of field \mathcal{F} and the distance transform is defined to be the best estimate for the center of the shape. Figure 5 (left) plots the distance and standard deviation of the estimated center to the real center of the circle averaged over 1000 trials for different levels of noise.

Our medial features provide significantly higher robustness against noise than the symmetry point detection using the ridges on D_T . This is due to the fact that they are based on an anisotropic diffusion process that suppresses spurious noise. It could be argued that a preprocessing step such as edge map erosion would increase the performance of the symmetry point detection based on the distance transform. However, this would also hold true for our SCIPs and we view the noisy input edge map as the result of such preprocessing steps which may still not be able to produce a noise free image.

A common problem in computer vision is that interest points are often formed or influenced by intersections between the shape of interest and background clutter. Medial features are less sensitive against clutter than corner interest points which is illustrated in Figure 5 (right). As the vector field V_N on the inside of the tripod only depends on the edges of the tripod the medial features are not influenced by clutter. The SIFT interest points depend on local image structure close to the edges and consequently change positions as the background forms new corners with the shape of interest.

4.2. Experiment 2: Edge map reconstruction

Image encoding by expressing the image as a list of interest points and descriptors is an important but lossy step in

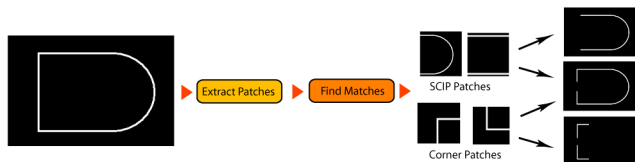


Figure 6. The edge map reconstruction pipeline. Patches are extracted from the edge map and compared to a codebook. The matching patches are rendered onto a canvas. Images on the right represent reconstruction from SCIP, corner based and mixed codebooks.

computer vision. We aim to quantify the portion of image information that is lost during this step. We investigate this using an edge map reconstruction task based on a codebook of visual words drawn from SCIPs and edge-based interest points (in our case SIFT keypoints).

The dataset consists of 50 images for training and 50 for testing taken from the the Berkeley Segmentation Dataset [29] which consists of natural photographs. In a first step we created a codebook of square edge map patches from the training image set. This codebook is obtained by extracting patches at SCIP and SIFT interest points using the estimated scales and orientations of both resulting in scale and rotation invariant patches. To reduce redundancies in the codebook we perform an agglomerative clustering using normalized cross correlation as a distance measure. We obtain the final codebooks by kMeans clustering where k is the fixed size of the particular codebook.

In the reconstruction phase we match our codebook entries to the interest points found in images from the test set. If the normalized cross correlation is above the threshold of 0.7 we render the stored edge map patch from the codebook onto a canvas. The reconstruction pipeline is depicted in Figure 6. As performance measure we used the Euclidean distance between original and reconstructed edge map. Figure 7 (left) shows the performance for different codebook sizes. For very small codebook sizes SIFT is better able

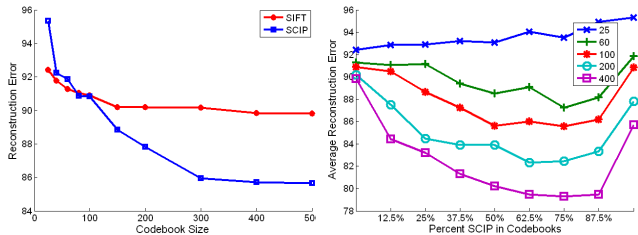


Figure 7. Edge map reconstruction performance. **Left:** Mean Euclidean distance of the original edge maps to the reconstructed edge maps for SIFT and SCIP as interest point detectors. **Right:** The reconstruction performance using mixtures of both codebooks. The legend denotes the size of the combined codebooks.

to reconstruct the edge maps. This is to be expected since SIFTs are edge-based which is initially preferable for the task. However, for practical codebook sizes reconstruction based on SCIP codebooks outperforms SIFT.

In a second experiment we evaluate the edge map reconstruction using mixed codebooks. We reconstruct the edge map from both kinds of interest points at different ratios to one another and fuse them by calculating the weighted average. The total size of the mixed codebook is kept constant and just the percentage of SCIP patches in it changes. No optimization of the combination is performed. The performances are shown in Figure 7 (right). The combination of shape centered and corner based interest points clearly outperform the codebooks from a single source. This indicates that medial features and SIFT encode different kinds of information and that a combination of both might be useful for computer vision algorithms. Since they are shape centered, SCIPs also offer a direct way to encode the color of the shape. This fact can be used to reconstruct the original image (instead of just the edge map) or augment feature vectors to avoid *e.g.* tracking ambiguities.

In addition to the edge reconstruction task we provide further evidence that SIFT and SCIP denote two distinct information sources. We apply a multidimensional scaling (MDS) algorithm to patch-based codebooks sampled at the two kinds of interest points. Figure 8 shows the first two dimensions of the embedding computed by the MDS with an Euclidean distance measure. The distinct separation of the two classes demonstrates that there is little overlap in the encoded information between shape-centered and corner-based interest points.

4.3. Experiment 3: Multiclass Classification

We now evaluate the information available at SIFTs (*i.e.* its descriptor) vs. the available information at SCIPs after grouping as described in Section 3.3 (which are for this purpose only the SIFT descriptors at the linked SIFT interest points). To this end we extract SIFTs and link them to our medial features on images from the StreetScenes

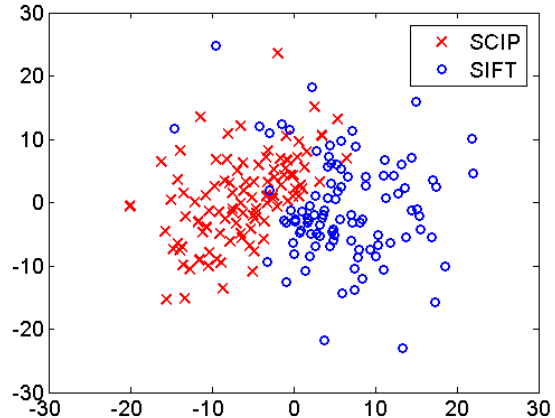


Figure 8. MDS clustering: The two code-books are located at two distinct clusters in the latent MDS space indicating the complementary nature of edge based and shape centered interest points.

database [30]. This database contains more than 3500 images taken in Boston with associated labels for eight classes (Pedestrian, Bicycle, Car, Street, Sidewalk, Building, Sky, Tree). An example image from the database with linked interest points are shown in Figure 9.

To demonstrate the power of the feature grouping by medial features we perform a recognition task with eight classes. We randomly sample 200 SIFT interest points and 200 SCIPs with the connected SIFT interest points for each of the 8 classes from the images resulting in 1600 training samples for each feature type. In the next step we use a kNN classifier (with $k = 5$) for classification. The simplicity of the classifier is chosen deliberately to minimize any possible effects of training and parameter estimation on the results. The distance measure between the SIFTs is the Euclidean distance. The distance measure for the grouped features is slightly more complex since it requires us to compare two graphs with different numbers of connected SIFTs. Using a codebook of visual words approach is not applicable in this case because there are on average only 46 connected SIFTs per SCIP. This would lead to feature vectors with only few non-zero entries making distance judgments between the graphs noisy and unreliable.

Thus, we perform distance judgments for grouped features using the standard SIFT matching scheme. We compare two graphs by pairwise calculation of the distances between all corner interest points. Then, we count how many matches there are between the smaller of the two graphs to the bigger of the two graphs. We accept two SIFTs from the two graphs as matches if and only if the ratio of the distance between the best and the second best match is larger than 0.8. This scheme has been proposed by David Lowe [3] and avoids having to set a threshold for matches in Euclidean distance which is often difficult and produces

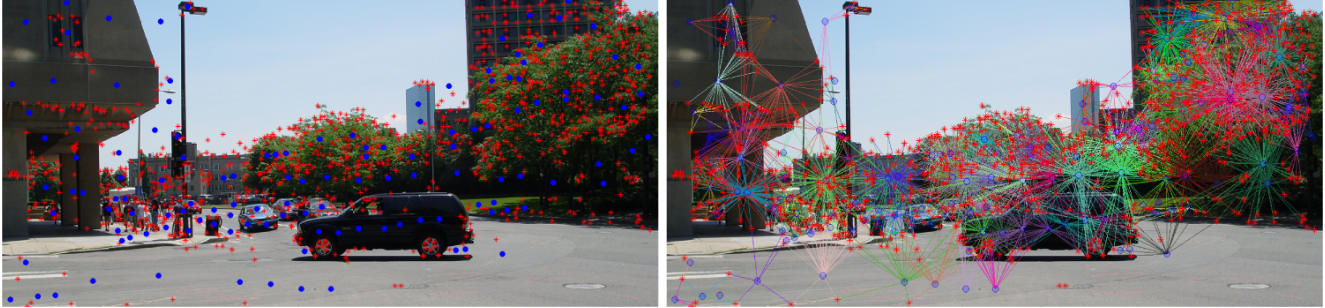


Figure 9. Downscaled image from the StreetScenes database with SIFT (red) and SCIP (blue) features. The right shows an overlay of the groupings from the SIFTs to the SCIPs.

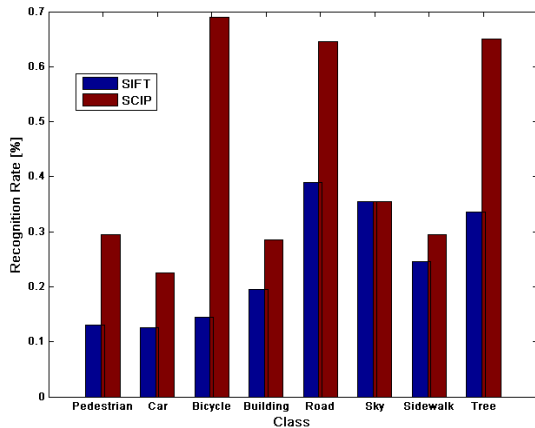


Figure 10. The classification rates for SIFTs and SCIPs. SCIPs were classified based on grouped SIFTs. The mean recognition rates across all classes are 24% for SIFT and 43% for SCIP. The baseline is 12.5%

unreliable results. With the resulting distance measure we can apply the kNN algorithm to classify SCIPs based solely on the grouped SIFTs. Our test set contained 200 examples for each of the classes for each type of interest point. The classification rates for each class are shown in Figure 10. The classification rates for the foreground classes 'car' and 'pedestrian' cannot keep up with state of the art detection schemes but this is to be expected since we do not propose a specific object detection framework. We used this multi-class classification task to demonstrate the power of the descriptors obtained by our grouping approach.

We want to stress that for both types of interest points only the SIFT descriptors are available meaning that both have to rely on the same type of information. Yet, by linking the SIFT information in a purely feed-forward manner at the shape-centered interest points we gain a more powerful description of the local image structure. Further increases in performance would be possible by augmenting the descriptor at the medial features by local color, shape and texture information or by introducing geometric constraints in the

distance measure between two graphs.

5. Conclusion

In this paper we have presented a robust, shape centered features based on GVF fields we call SCIP. We use a flux flow detection operation to determine points of high symmetry in the image. By being centered on shapes the new interest points provide more stability against clutter than corner based interest points. They are based on a diffusion process which makes them more stable against noise than standard medial features which are computed as ridges of the distance transformation. We have shown that SCIP encode relevant shape information of the image within an edge map reconstruction task. A combination of our medial features and corner based interest points performs better than each feature set alone which suggests that they encode independent information about the image. In Section 4.3 we have shown their usefulness for feature grouping in a multi-class recognition task on a dataset containing natural images.

Acknowledgements

This work was supported by EU-Project BACS FP6-IST-027140, DFG Perceptual Graphics and the Max Planck Society.

References

- [1] Xu, C., Prince, J.: Snakes, shapes, and gradient vector flow. *IEEE Transactions on Image Processing* **7** (1998) 359–369 [2](#)
- [2] Engel, D., Curio, C.: Scale-invariant medial features based on gradient vector flow fields. In: *ICPR*. (2008) 1–4 [2, 3](#)
- [3] Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision* **60**(2) (2004) 91–110 [2, 6](#)

- [4] Kimia, B.B.: The role of propagation and medial geometry in human vision. *Journal of Physiology* **97** (2003) 155–190 [2](#)
- [5] Tarr, M.J., Bülthoff, H.H.: Image-based object recognition in man, monkey and machine. *Cognition* **67**(1-2) (1998) 1–20 [2](#)
- [6] Lee, T., Mumford, D., Romero, R., Lamme, V.: The role of the primary visual cortex in higher level vision. *Vision Research* **38** (1998) 2429–2454 [2](#)
- [7] Giese, M.A., Poggio, T.: Neural mechanisms for the recognition of biological movements. *Nature Review Neuroscience* **4**(3) (2003) 179–192 [2](#)
- [8] Serre, T., Wolf, L., Poggio, T.: Object recognition with features inspired by visual cortex. In: *CVPR*. Volume 2. (2005) 994–1000 [2](#)
- [9] Riesenhuber, M., Poggio, T.: Hierarchical models of object recognition in cortex. *Nature Neuroscience* (1999) 1019–1025 [2](#)
- [10] Bileschi, S., Wolf, L.: Image representations beyond histograms of gradients: The role of gestalt descriptors. In: *IEEE Conference on Computer Vision and Pattern Recognition*. (2007) 1–8 [2](#)
- [11] Zhu, S.: Embedding gestalt laws in markov random fields. *IEEE transactions on pattern analysis and machine intelligence* **21**(11) (1999) 1170–1187 [2](#)
- [12] Zhu, S., Yuille, A.: Forms: A flexible object recognition and modelling system. *International Journal of Computer Vision* **20**(3) (Jan 1996) 187–212 [2](#)
- [13] Stahl, J., Wang, S.: Edge grouping combining boundary and region information. *IEEE Transactions on Image Processing* **16** (October 2007) 2590–2606 [2](#)
- [14] Ren, X., Fowlkes, C.C., Malik, J.: Learning probabilistic models for contour completion in natural images. *Int. J. Comput. Vision* **77**(1-3) (2008) 47–63 [2](#)
- [15] Guo, C., Zhu, S., Wu, Y.: Modeling visual patterns by integrating descriptive and generative methods. *International Journal of Computer Vision* **53**(1) (2003) 5–29 [2](#)
- [16] Guo, C., Zhu, S., Wu, Y.: Towards a mathematical theory of primal sketch and sketchability. *Proc. ICCV* **1228** (2003) [2](#)
- [17] Siddiqi, K., Bouix, S., Tannenbaum, A., Zucker, S.: The hamilton-jacobi skeleton. In: *International Conference on Computer Vision*. (1999) 828–834 [2](#)
- [18] Dimitrov, P., Damon, J.N., Siddiqi, K.: Flux invariants for shape. In *Computer Vision and Pattern Recognition* **1** (2003) 835–841 [2](#)
- [19] Goh, W.B., Chan, K.Y.: Shape description using gradient vector field histograms. In: *Scale Space Methods in Computer Vision*, Springer-Verlag (2003) 1611–3349 [2](#)
- [20] Goha, W.B., Chanb, K.Y.: The multiresolution gradient vector field skeleton. *Pattern Recognition* **40** (2007) 1255–1269 [2](#)
- [21] Lindeberg, T.: Edge detection and ridge detection with automatic scale selection. *International Journal of Computer Vision* **30**(2) (1998) 117–154 [2](#)
- [22] Lu, C., Pizer, S., Joshi, S., Jeong, J.: Statistical multi-object shape models. *International Journal of Computer Vision* **75** (2007) 387–404 [2](#)
- [23] Kadir, T., Brady, M.: Saliency, scale and image description. *Int. J. Comput. Vision* **45**(2) (2001) 83–105 [2](#)
- [24] Fergus, R., Perona, P., Zisserman, A.: Object class recognition by unsupervised scale-invariant learning. In: *Conf. on Computer Vision and Pattern Recognition*. Volume 2. (June 2003) 264–271 [2](#)
- [25] Toshev, A., Shi, J., Daniilidis, K.: Image matching via saliency region correspondences. In: *CVPR*. (2007) [2](#)
- [26] Tu, Z., Chen, X., Yuille, A., Zhu, S.: Image parsing: Unifying segmentation, detection, and recognition. *IJCV* **63**(2) (2005) 113–140 [2](#)
- [27] Zheng, S., Tu, Z., Yuille, A.: Detecting object boundaries using low-, mid-, and high-level information. *Proc. of CVPR* (2007) [2](#)
- [28] Pizer, S., Siddiqi, K., Szekeley, G., Zucker, S.: Multi-scale medial loci and their properties. *Int. J. of Computer Vision* **55**(2-3) (2003) 155–179 [3](#)
- [29] Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: *ICCV*. Volume 2. (July 2001) 416–423 [5](#)
- [30] Bileschi, S.M.: Streetscenes: towards scene understanding in still images. PhD thesis, Massachusetts Institute of Technology (2006) [6](#)