

# Curvature Based Robust Descriptors

Farlin Mohideen  
farlin.mohideen@anu.edu.au  
Ranga Rodrigo  
ranga@ent.mrt.ac.lk

Australian National University  
Department of Electronic and  
Telecommunication,  
University of Moratuwa

---

## Abstract

Feature descriptors have enabled feature matching under varying imaging conditions, while mostly being backed by experimental evidence. In addition to imposing some restrictions in imaging conditions needed to ensure matching, extending the existing descriptors is not straightforward due to the lack of sound mathematical bases. In this work, by using a surface bending versus shape histogram based on the principal curvatures, we are able to produce a descriptor which is not sensitive to the errors in dominant orientation assignment. Experimental evaluations show that our descriptor outperforms existing descriptors in the areas of viewpoint, rotation, scale, zoom, lighting and compression changes, with the exception of resilience to blur. Further, we apply this descriptor for accuracy demanding applications such as homography estimation and pose estimation. The experimental results show significant improvements in estimated homography and pose in terms of residual error and Sampson distance respectively.

## 1 Introduction

The task of finding feature correspondences between images is a part of many computer vision applications. The main advantage of this method is that it uses a small pixel patch around the interest point to produce the descriptor, unlike region based matching techniques. This ensures that even if the second image has less in common corresponding to the first image, correspondences still could be found. Feature point description is basically the task of describing the region around an interest point which can be a corner or blob in an invariant or partially invariant manner to viewpoint, rotation, scale, zoom, lighting, compression and blur, in a much lower dimension space, with respect to the pixel patch around the interest point [14]. Computation complexity is also a great concern when it comes to descriptor's practical usability [23]. Applications of feature descriptors in the literature include wide baseline matching [18], object recognition [3, 10], texture recognition [8], image retrieval [12],[19], robot localization [20], video data mining [22] and object category recognition [9]. Also, in tracking applications, despite the view point changes they can be used for motion estimation. Feature descriptors, in this context, have made a remarkable contribution to computer vision.

Kenney *et al.* [9] describe an axiomatic approach for key point detection which provides a rigorous explanation of the notion of detection. Unfortunately, descriptor theory is not as

complete as detector theory, and this has made lot of researchers follow heuristic methods to come up with better descriptors. Many new descriptors have been formed using different metrics and methodologies.

In this paper, we describe a new descriptor based on curvature with mathematically sound argument. We justify that curvature is a more suitable metric, using the property that it does not depend on the local coordinate frame changes and produces the same value for any coordinate frame due to plane rotation and translation. SIFT [10] descriptor metrics, gradient magnitude vs gradient angle histograms, are indicated to be inspired from biological vision. Our choice of the metrics have a more mathematically sound justification, where we consider matching as a 2D-surface (patch taken around the key-point) matching problem. We also demonstrate that our method is more resilient to noise than SIFT and SURF. Furthermore, we explain why spatial patches are needed, establish and prove a better way to get the patch around the key point, and explain why a damping window is needed. In this manner, we give mathematically sound justifications to the steps used in our descriptor. The main contributions of this work are

- introducing a descriptor with computation complexity on par with SIFT, while outperforming other descriptors,
- formulating our descriptor on mathematical reasoning and thus justifying its advantages, and
- empirically showing the superior performance of this descriptor for homography estimation and pose estimation .

## 2 Related Work

A wide variety of detectors and descriptors [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15] have been proposed in the literature, and detailed comparisons have been performed [16, 17]. According to these, state-of-the-art descriptors which do not use machine learning include descriptors such as SIFT [10], SURF [11], and GLOH [12]. Techniques that require machine learning include Ferns [15] and Robust Feature Matching [13]. These may not be well suited for all applications as the learning phase takes time and are mostly suitable for controlled environments. Thus, we choose to work on non-machine-learning type features. Based on SIFT, Moreno *et al.* [18] describe a method to improve the SIFT descriptor in terms of matching performance and Grabner *et al.* [9] describe an efficient way to compute the SIFT descriptor using integral images [14]. PCA-SIFT [6] claims a more distinctive SIFT-like descriptor, but was later shown to be not as robust as SIFT [15] and it is computationally expensive. SURF [11] is inspired by SIFT, and performs as good as SIFT for rotation and image blur, and is several times faster. Still, SIFT is better when it comes to viewpoint, illumination, scale and zoom [10]. Miklojczyk *et al.* [19] describe a way to find affine normalized regions, but, this is known to be computationally expensive and it is not as good as viewpoint sampling [16]. Our descriptor was inspired by SIFT and SURF, but its formulation steps are mathematically sound. Furthermore, our descriptor does not employ the technique of viewpoint sampling proposed by Morel and Yu [16], but rather focuses on moderate projective invariance under the affine assumption.

## 3 Methodology

Stages of feature matching involve finding interest points in a scale space [9], usually a blob or corner like structure, finding the dominant orientation around the keypoint [10] and, descriptor formation by utilizing the region around the keypoint w.r.t. the dominant orientation [10].

### 3.1 Interest Point Detection

Most widely used detectors include Harris [9], Hessian Laplace [10] and determinant of Hessian. Linderberg [9] proposed a way for automatic scale selection, which made it possible to detect interest points stable both in location and scale. Since good interest point detectors that are robust to scale changes are already available, we focus on the descriptor. For our test cases we used DoG [10] for all tested feature descriptors to make the comparison straightforward.

### 3.2 Interest Point Description

Interest point description has two steps: orientation assignment, and description [14]. Orientation assignment is the stage in which the pixel patch around the interest point in the image scale-space [10] is chosen with respect to a dominant orientation, invariant under many deformations. It is the key step for rotational invariance of the descriptor.

#### 3.2.1 Characterization of a Good Descriptor

An invariant dominant orientation for the grid is needed to make the descriptor rotationally invariant in the presence of more than one spatial patch for descriptor formation (unless the spatial patches are polar rings and a rotationally invariant metric such as curvature is used). Many methods to find the dominant orientation has been proposed including orientation of the largest eigenvector in Harris [9], maxima in the edge orientation histogram [10], gradient direction at a very coarse scale and maxima of Haar wavelet response for a rotating arc [10]. No matter which statistic or method is used to compute the dominant orientation, we believe, that a proper descriptor design should consider the possibility of misorientation, as the statistic used is not proven explicitly to be invariant under patch deformations. Misorientation makes oriented spatial patches gain new pixels and loose pixels near the corners of each spatial patch. This happens mostly to spatial patches near the periphery of the patch. Thus, to reduce the influence of spatial patches near the periphery a damping window is used (Gaussian window in SIFT), and also this improves the resilience to mis-registration. However this damping window also serves another purpose. Before coming to that point, let us investigate what kind of a grid size should be chosen around the patch. SIFT uses a  $16 \times 16$  pixel patch. We show that the grid width should be proportional to the scale of the key point. In a signal processing context it is shown that the effective time  $T_d$  bandwidth  $B_w$  product is lower bounded:

$$T_d \cdot B_w \geq \frac{1}{2}. \quad (1)$$

$$T_{d\min} \cdot B_{w\max} = \text{constant}. \quad (2)$$

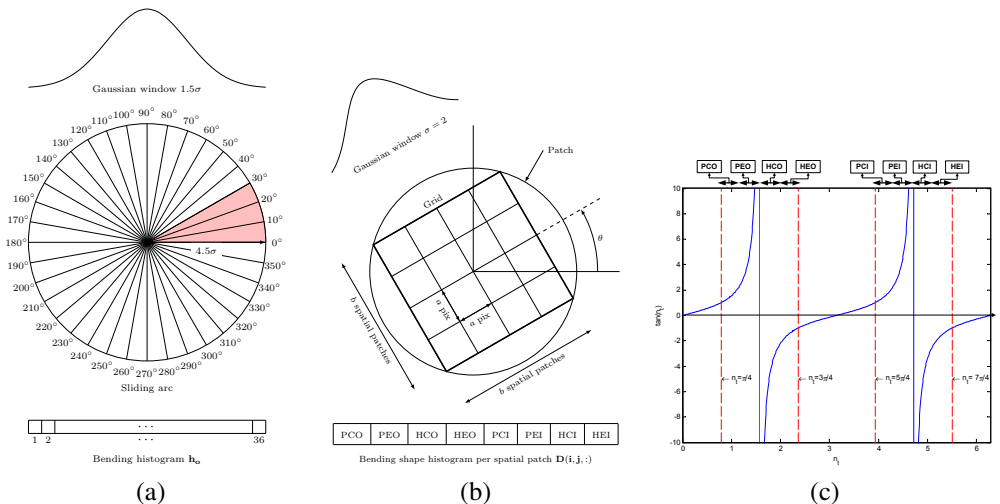


Figure 1: An image patch divided into 36 arc-windows. A grid superimposed on an image patch, and a grid divided into spatial patches. A Gaussian damping window is overlaid on each patch. (a) Bending histogram for dominant orientation. (b) Descriptor histogram. (c) Eight shape classifications and corresponding tan-curve fitted values.  $a = 3\sigma$  is the width of a spatial patch. Descriptor grid width  $b = 4$ . Number of classification bins  $c = 8$ . The gaussian damping window  $g$  for descriptor is of variance  $\sigma_0 = 2$ . To meet these requirements and considering extra patches needed in distributing bending among adjacent bins we need to consider a patch of radius  $w = \sqrt{2}a(b+1)/2 + 0.5$ .

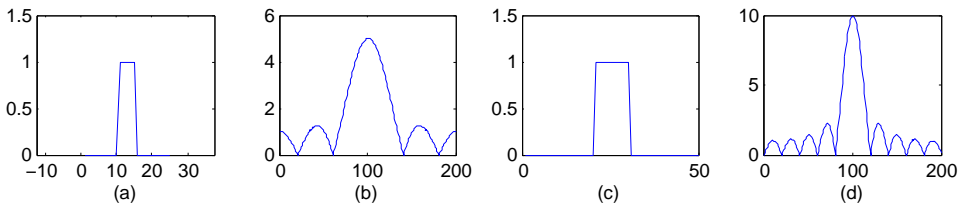


Figure 2: Corresponding signals for the same shape under half the bandwidth of the other. (a) First signal. (b) Frequency response of first signal. (c) Second signal with double the width of the first signal. (d) Frequency response of the second signal.

$$\frac{T_{\text{dmin}1}}{T_{\text{dmin}2}} = \frac{B_{\text{wmax}2}}{B_{\text{wmax}1}} = \frac{\sigma_1}{\sigma_2}. \quad (3)$$

The signal's observed duration should be increased to represent the same shape in a much lower bandwidth (1). Figure 2 shows signals with the same shape but different bandwidths. Thus we employ the constraint (2) which can be deduced to (3), since the maximum bandwidth of a signal is upper bounded by the Gaussian convolution kernel's bandwidth corresponding to its scale, which, in turn, is inversely proportional to  $\sigma$ . Furthermore, the lower bound of the effective duration ensures that the grid captures the most needful area around the key point structure. However, it does not guarantee that the window size will always en-

capsulate exactly the structure of interest; it may capture more than intended. The third use of the damping window is that it minimizes the influence of this extra captured region. This can also be thought as minimizing spectral leakage in the frequency domain. Thus, in our descriptor, we use a grid, with width proportional to the scale, along with a damping window.

Furthermore misorientations that occur due to patch deformation and mis-registration can be handled by using a rotationally invariant metric, such as principal curvatures. SIFT uses gradient orientation histogram [10], while SURF uses Haar wavelet responses [11] as the descriptor metric. In SIFT, significant misorientation causes histogram distortion as the binning position is dependent on gradient angle w.r.t. the dominant orientation, thus making binning positions change when an error occurs in the dominant orientation computation. In SURF, Haar wavelet response error occurs under misorientation as Haar wavelet responses are not rotationally invariant. In order to counter this we propose to use a histogram metric which is truly invariant under rotation and translation: Principal curvatures are invariant under rotation and translation. Furthermore, principal curvatures describe the surface nature at a point, thus being more representative of the surface. The surface itself can be classified according to the amount of bending and shape (Figure 3). We use these important properties of curvatures to formulate a meaningful descriptor.

### 3.2.2 Orientation Assignment

In differential geometry the eigenvalues of Hessian matrix  $H$  are the principal curvatures  $\lambda_{\max}, \lambda_{\min}$  of a surface  $I(x, y)$  at any given point. Let  $\vec{p}$  be any point in patch  $S$ . We introduce a metric the amount of bending  $m(\vec{p})$  based on rotationally invariant principle curvatures as below

$$H = \begin{bmatrix} I_{xx} & I_{xy} \\ I_{xy} & I_{yy} \end{bmatrix}, \quad H\vec{v} = \lambda\vec{v}, \quad m(\vec{p}) = \sqrt{\lambda_{\max}^2 + \lambda_{\min}^2}. \quad (4)$$

We propose to represent the dominant orientation by finding where the maximum bending of the surface occurs in a sliding arc-window of  $30^\circ$  (Figure 1a). We compute this statistic in a patch  $S$  with center  $(x_0, y_0)$  and with radius proportional to scale for the aforementioned reasons (Figure 1a). Let  $\vec{p}$  be any point in patch  $S$  and  $f(\vec{p})$  be the polar angle of  $\vec{p}$  w.r.t  $(x_0, y_0)$  as shown below.

$$\vec{p} = \begin{bmatrix} x \\ y \end{bmatrix}, \quad S = \{\vec{p} | (x - x_0)^2 + (y - y_0)^2 \leq 3 \times 1.5\sigma\}. \quad (5)$$

$$f(\vec{p}) = \tan^{-1}(y - y_0, x - x_0), \quad \theta_0 = 10^\circ, \quad N = 360/\theta_0 = 36. \quad (6)$$

We compute the histogram  $h_o$  of bending of the surface  $m(\vec{p})$  (4), vs. polar angle of the pixel  $f(\vec{p})$  w.r.t. keypoint  $(x_0, y_0)$ , multiplied by a Gaussian damping window  $g(r, 1.5\sigma)$  where  $\alpha, \beta, r$  are used for linear interpolation among adjacent bins, as represented by

$$\alpha = \frac{f(\vec{p})}{\theta_0}, \quad \beta = \text{floor}(\alpha - 0.5), \quad r = (\alpha - \beta - 0.5). \quad (7)$$

$$h_o(i) = \sum_{\vec{p} \in S} \{m(\vec{p})g(r, 1.5\sigma) [\delta(i - (\beta + N) \bmod N)(1 - r) + \delta(i - (\beta + 1) \bmod N)r]\} (8)$$

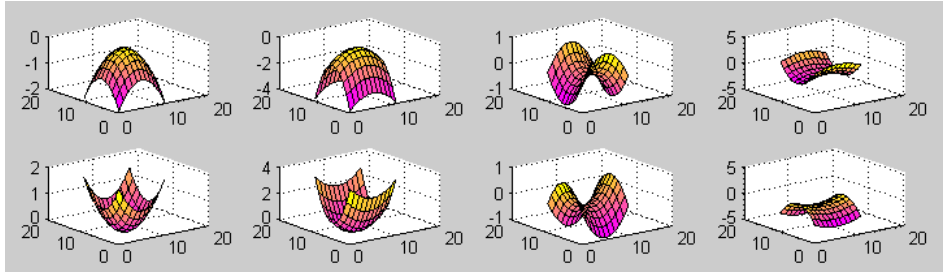


Figure 3: Shape classifications according to parabola (P), hyperbolae (H), corner (C), edge (E), outward (O) and inward (I). (a) PCO (b) PEO (c) HCO (d) HEO (e) PCI (f) PEI (g)HCI (h)HEI

Thus, the bending histogram  $h_o(i)$  is computed where each bin corresponds to the  $10^\circ$ -arc's total bending of the surface under the influence of the Gaussian. Binning is done by distributing values among adjacent bins by trilinear interpolation (7),(8). Responses for the  $30^\circ$ -arc with  $10^\circ$  sliding is found by creating the final histogram  $\hat{h}_o(i)$  by summing three adjacent bins of  $h_o(i)$  by

$$\hat{h}_o(i) = \sum_{j=1, \dots, 3} [h_o((i+j) \bmod N)], \quad i = 1, \dots, N. \quad (9)$$

In  $\hat{h}_o(i)$  orientation values are found for the highest peak and those above 75% of the highest [11, 12], followed by interpolation as in SIFT [13]. Each dominant orientation is used to find descriptors; as in SIFT, one keypoint may have multiple descriptors [14]. In summary, a good descriptor is preferably patch based, the grid width being proportional to the scale, resilient to misorientation due to relying on a metric like curvature which characterizes the shape of the surface at any point.

### 3.2.3 Feature Description

There are four steps in our feature description: (1) Computing the rotated, normalized spatial patch coordinate frame (2) Surface classification for each patch (3) Generating the descriptor vector and (4) Normalization. We use a  $4 \times 4$  spatial patch grid for our descriptor, with width  $12\sigma$  (as shown previously grid width should be proportional to scale) rotated to the dominant orientation which is handled by normalized rotated coordinates  $N_x, N_y$  as shown below where  $\theta$  is the computed dominant orientation,  $(x_0, y_0)$  is the key-point and  $(N_x, N_y)$  is the spatial patch location for descriptor binning.

$$\begin{bmatrix} N_x \\ N_y \end{bmatrix} = \frac{1}{a} \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix} \quad (10)$$

For each spatial patch, we create a surface bending  $m(\vec{p})$  vs shape histogram  $D(i, j, :)$  (must see Figure 1b,) based on the eight classifications of the surface (must see Figure 3).

We classify the amount of surface bending according to shapes (Table 1) based on the ratio of principal curvatures. Each spatial patch produces an eight-element descriptor and all 16 spatial patches produce a 128D descriptor similar to SIFT. Principal curvature max and min are chosen in an absolute sense to make sure that any point will fall in to one of the classifications as shown below.

$$\lambda_{\max} = \begin{cases} \lambda_1, & |\lambda_1| > |\lambda_2|, \\ \lambda_2, & \text{otherwise.} \end{cases} \quad \lambda_{\min} = \begin{cases} \lambda_1, & |\lambda_1| \leq |\lambda_2|, \\ \lambda_2, & \text{otherwise.} \end{cases} \quad (11)$$

Table 1: Shape classifications at each pixel: paraboloid (P), hyperboloid (H), edge (E), corner (C), outward (O), inward (I),  $K = \lambda_{\max}\lambda_{\min}$  and  $J = \lambda_{\max}/\lambda_{\min}$ .

| Metric | P       | H       | C            | E            | I                      | O                       |
|--------|---------|---------|--------------|--------------|------------------------|-------------------------|
| $K$    | $K > 0$ | $K < 0$ | -            | -            | -                      | -                       |
| $J$    | -       | -       | $ J  < 2.41$ | $ J  > 2.41$ | -                      | -                       |
| $n_t$  | -       | -       | -            | -            | $[\pi/4 \quad 3\pi/4]$ | $[5\pi/4 \quad 7\pi/4]$ |

Classification of bins is done in continuous linear space for efficiency. As the interval of curvature ratios that correspond to edges is far bigger than for corners, taking the ratio of principle curvatures does not provide a smoothly varying metric from edge to corner , thus we fit the curvature ratio to a tan curve considering  $\{\lambda_{\max}\} \times \{\lambda_{\min}\}$  cartesian space (Figure 1c) as shown below

$$n_t = \tan^{-1}(\lambda_{\max}, \lambda_{\min}). \quad (12)$$

Now the classification problem is in linear space for binning. As we use  $\lambda_{\max}, \lambda_{\min}$  with algebraic sign, but max, min are chosen in an absolute sense (11) half the linear space values retrieved by (12) are not occupied, and only a portion of  $180^\circ$  is occupied (see Figure 1c). Thus, we will not get our intended classification. To prevent this, we normalize  $n_t$  to be classified into eight bins ( $c = 8$ ) spanning  $360^\circ$  by

$$N_t = \begin{cases} \frac{c(n_t - \pi/4)}{\pi}, & n_t \leq 3\pi/4, \\ c/2 + \frac{c(n_t - 5\pi/4)}{\pi}, & n_t \leq 7\pi/4. \end{cases} \quad (13)$$

Normalized spatial coordinates for spatial binning could be obtained from (10) while  $N_t$  provides the normalized shape bin. Now, according to  $N_x, N_y, N_t$ , the amount of bending  $m(\vec{p})$  multiplied by the Gaussian is binned to the descriptor histogram  $D(i, j, k)$  by distributing among adjacent bins by trilinear interpolation considering that spatial bins are not circularly buffered and classification bins are circularly buffered. The amount of bending  $m(\vec{p})$  multiplied by the gaussian damping window value  $g(r, \sigma_0)$  is binned to descriptor histogram  $D(i, j, k)$  at location  $(N_x, N_y, N_t)$  by distributing among adjacent bins by the following trilinear interpolation equations

$$\beta_x = \text{floor}(N_x - 0.5), \quad \beta_y = \text{floor}(N_y - 0.5), \quad \beta_t = \text{floor}(N_t). \quad (14)$$

$$r_x = N_x - (\beta_x + 0.5), \quad r_y = N_y - (\beta_y + 0.5), \quad r_t = N_t - \beta_t. \quad (15)$$

$$H(i, r_a, \beta_a) = \begin{cases} 1 - r_a, & i = \beta_a, \\ r_a, & i = \beta_a + 1, \\ 0, & \text{otherwise.} \end{cases} \quad a = x, y \text{ or } t. \quad (16)$$

Let  $D(i, j, k)$  be the  $4 \times 4 \times 8$  descriptor where  $i = 1, \dots, 4$ ,  $j = 1, \dots, 4$ ,  $k = 1, \dots, 8$ , which is a histogram formed as shown below

$$D(i, j, k) = \sum_{\vec{p} \in P} \{m(\vec{p}) g(r, \sigma_0) H(i, r_x, \beta_x) H(j, r_y, \beta_y) H(k, r_t, \beta_t)\}. \quad (17)$$

where  $P$  is the patch with radius  $w$  used to compute the descriptor (see Figure 1.(b)) and  $H(i, r_x, \beta_x) H(j, r_y, \beta_y) H(k, r_t, \beta_t)$  represent the distribution of the term  $m(\vec{p}) g(r, \sigma_0)$  (gaussian weighted metric) among adjacent bins with trilinear interpolation. The descriptor  $D(i, j, k)$  is normalized and values greater than 0.1 are clipped and re-normalized for robustness to illumination changes.

## 4 Experimental Results

### 4.1 ROC Curves

We evaluate the performance of our descriptor in terms of viewpoint, zoom and rotation, lighting, compression and blur performance using the Oxford image database [14] which has structured (S) and textured (T) images. Based on the ground-truth homographies available with the database, we plotted ROC curves, the number of true positives versus false alarm rate. It is reasonable to use the number of true positives for comparison as we used the same detector (DoG) for both SIFT and CUR. In all situations, except blur, our descriptor outperforms SIFT (see Figures 4 and 5). Although the gradient profile is less affected by blur, the patch surface shape is significantly affected, which causes considerable curvature distortions. Apparently, this is the reason why our descriptor does not perform as well as SIFT in the presence of blur.

Table 2 approximately compares GLOH, SURF and CUR with SIFT in terms of recall. Our descriptor, CUR, outperforms all three descriptors except in the case of blur and zoom and rotation for a structured scene. SURF performs best in presence of blur and it performs as well as CUR for zoom and rotation.

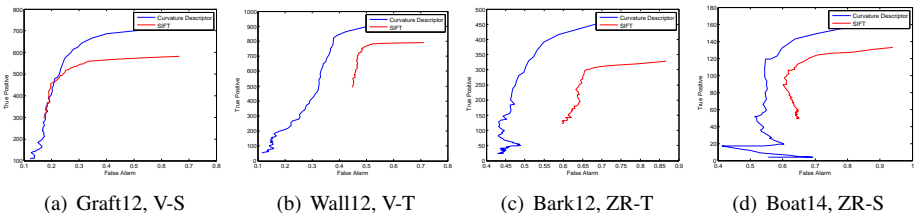


Figure 4: True positives vs false alarm rate curves in the presence of viewpoint (V) change ( $20^\circ$ – $30^\circ$ ), and zoom and rotation (ZR) ( $30^\circ$ ). S: structured T: textured

### 4.2 Homography Estimation

We evaluate homography estimation in terms of residual error computed by  $(\|x' - Hx\| + \|x - H^{-1}x'\|)/2$  where  $H$  is the estimated homography and  $(x, x')$  is a correspondence obtained by feature matching. Thus this is a measure that checks how well the



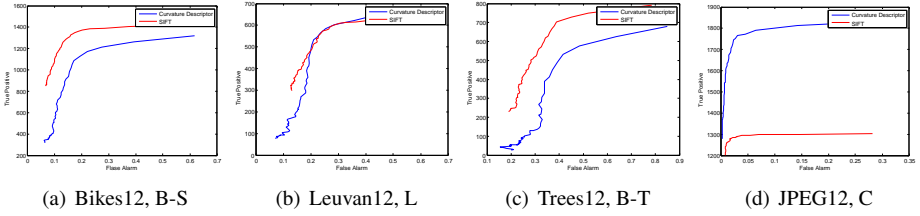


Figure 5: True positives vs false alarm rate curves in the presence of image blur (B), lighting (L), and compression changes (C) changes. S: structured, T: textured

Table 2: Approximate improvement and degradation of each descriptor w.r.t. SIFT. +: 0.1 recall improvement - : 0.1 recall degradation 0: a tie. Viewpoint (V), Zoom and Rotation (ZR), Blur (B), Lighting (L), Compression (C), Structured scene(S) and Textured scene (T).

| Descriptor | V-S | V-T   | ZR-S | ZR-T | B-S | B-T   | L  | C   |
|------------|-----|-------|------|------|-----|-------|----|-----|
| GLOH       | +   | -     | +    | 0    | 0   | -     | 0  | 0   |
| SURF       | -   | - - - | +++  | ++   | +   | +     | -- | --  |
| CUR        | ++  | +     | ++   | +++  | -   | - - - | 0  | +++ |

descriptor finds correspondences that are compatible with the estimated homography, which is a reasonable way of comparing when absolute ground truth comparison is meaningless. Figure 6.a shows the histogram of number of inlier key-points vs residual error for homography estimation for SIFT and CUR.

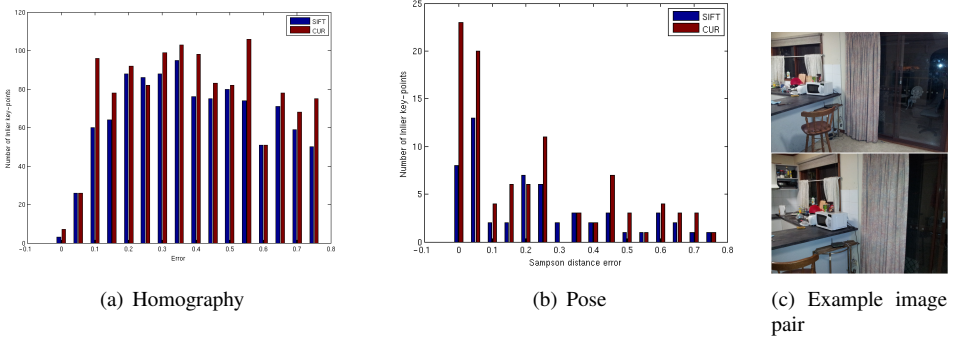


Figure 6: Number of inlier key-points vs residual error for homography estimation and number of inlier key-points vs Sampson distance for pose estimation.

### 4.3 Pose Estimation

We evaluated pose estimation in terms of first-order geometric error (Sampson distance) by  $\frac{(x'^T F x)^2}{(F x)_1^2 + (F x)_2^2 + (F^T x')_1^2 + (F^T x')_2^2}$ . We represent the accuracy by a histogram of number of inlier key-points vs their Sampson distance as shown by Figure 6.b.

## 5 Conclusion

In this paper we presented a novel descriptor based on the amount of surface bending vs shape histogram w.r.t. a dominant orientation, which is computed using the maximum amount of bending in an arc-window. Our formulation of the descriptor was based on the characteristics needed in a good descriptor, namely, the use of spatial patches, use of orientation, resilience to misorientation, grid width being proportional to scale, use of a damping window, and a rotationally invariant metric such as the curvature. Experimental results showed that our descriptor, CUR, outperforms all three descriptors SIFT, SURF, and GLOH in all situations, except in the case of blur and zoom and rotation for a structured scene. Also the experiments on practical real world applications such as homography estimation and pose estimation show significant improvements.

## References

- [1] H. Bay, T. Tuytelaars, and L.V. Gool. Surf: Speeded up robust features. 3951–2006: 404–417, May 2006.
- [2] G. Dorkó and C. Schmid. Selection of scale-invariant parts for object class recognition. In *IEEE International Conference on Computer Vision*, pages 634–640, Montbonnot, France, April 2003.
- [3] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 264–271, Oxford, the UK, June 2003.
- [4] M. Grabner, H. Grabner, and H. Bischof. Fast approximated SIFT. In *Asian Conference on Computer Vision*, pages 918–927, Hyderabad, India, 2006.
- [5] Chris Harris and Mike J. Stephens. A combined corner and edge detector. In *Proceedings of the 4th Alvey Vision Conference*, pages 147–152, Manchester, UK, August 1988.
- [6] Yan Ke and Rahul Sukthankar. PCA-SIFT: A more distinctive representation for local image descriptors. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 506–513, Washington, DC, June–July 2004.
- [7] C. S. Kenney, M. Zuliani, and B. S. Manjunath. An axiomatic approach to corner detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 191–197, San Diego, CA, June 2005.
- [8] Lazebnik, S. Schmid, and C. Ponce. A sparse texture representation using affine-invariant regions. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 319–324, Urbana, IL, June 2003.
- [9] Tony Lindeberg. Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):79–116, November 1998.
- [10] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):31–110, November 2004.

- [11] 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*, 65(1-2):43–72, November 2005.
- [12] Krystian Mikolajczyk and Cordelia Schmid. Indexing based on scale invariant interest points. In *International Conference on Computer Vision*, volume 1, pages 525–531, Montbonnot, France, July 2001.
- [13] Krystian Mikolajczyk and Cordelia Schmid. Scale and affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1):63–86, October 2004.
- [14] Krystian Mikolajczyk and Cordelia Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, October 2005.
- [15] Pierre Moreels and Pietro Perona. Evaluation of features detectors and descriptors based on 3-D objects. In *Proceedings of the Tenth IEEE International Conference on Computer Vision*, volume 1, pages 800–807, Washington, DC, October 2005.
- [16] J.M. Morel and G.S. Yu. ASIFT: A new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences*, 2(2):438–469, 2009.
- [17] P. Moreno, A. Bernardino, and S.J. Victor. Improving the SIFT descriptor with smooth derivative filters. *PRL*, 30(1):18–26, January 2009.
- [18] F. Schaffalitzky and A. Zisserman. Multi-view matching for unordered image sets, or “How do I organize my holiday snaps?”. In *European Conference on Computer Vision, Copenhagen, Denmark*, volume 1, pages 414–431. Springer-Verlag, 2002.
- [19] Cordelia Schmid and Roger Mohr. Local grayvalue invariants for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:530–535, 1997.
- [20] Stephen Se, David Lowe, and Jim Little. Global localization using distinctive visual features. In *International Conference on Intelligent Robots and Systems*, volume 1, page 226, City, ON, Canada, December 2002.
- [21] A Shokoufandeh, I Marsic, and S.J Dickinson. View-based object recognition using saliency maps. *Image and Vision Computing*, 17(5–6):445–460, April 1999.
- [22] Josef Sivic, Bryan C. Russel, Alexie A. Efros, Andrew Zisserman, and William T. Freeman. Discovering objects and their location in images. In *IEEE International Conference on Computer Vision*, pages 370–377, Beijing, China, October 2005.
- [23] Simon Taylor, Edward Rosten, and Tom Drummond. Robust feature matching in  $2.3\mu$  s. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 15–22, Miami, FL, June 2009.
- [24] Paul Viola and Michael J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, May 2004.
- [25] M. Özuysal, M. Calonder, V. Lepetit, and P. Fua. Fast keypoint recognition using random ferns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32:448, 2009.