

Nested Shape Descriptors

Jeffrey Byrne and Jianbo Shi
University of Pennsylvania, GRASP Lab
{jebyrne, jshi}@cis.upenn.edu

Abstract

In this paper, we propose a new family of binary local feature descriptors called nested shape descriptors. These descriptors are constructed by pooling oriented gradients over a large geometric structure called the *Hawaiian earring*, which is constructed with a nested correlation structure that enables a new robust local distance function called the *nesting distance*. This distance function is unique to the nested descriptor and provides robustness to outliers from order statistics. In this paper, we define the nested shape descriptor family and introduce a specific member called the *seed-of-life descriptor*. We perform a trade study to determine optimal descriptor parameters for the task of image matching. Finally, we evaluate performance compared to state-of-the-art local feature descriptors on the VGG-Affine image matching benchmark, showing significant performance gains. Our descriptor is the first binary descriptor to outperform SIFT on this benchmark.

1. Introduction

Local feature descriptors have emerged in the past ten years as the dominant representation for image matching. There exist standard benchmarks for performance evaluation [11, 12, 21], and a zoo of detectors and descriptors [1, 14, 4, 20, 8, 3, 17, 15]. introduced with the trend of faster and faster matching while maintaining approximately equivalent performance to SIFT [9]. Local feature descriptors have been successfully deployed for a wide range of image matching tasks including: stereo, optical flow, structure from motion, egomotion estimation, tracking, geolocation and mapping.

All existing local feature descriptors share a common performance tradeoff between support size and matching selectivity. It is well known that for the task of image matching, descriptors constructed with larger support outperform descriptors with smaller support [20, 8, 3, 17, 15]. Descriptors with large support are constructed with larger image patches that increase the uniqueness of a match and address the aperture problem, however there are diminishing

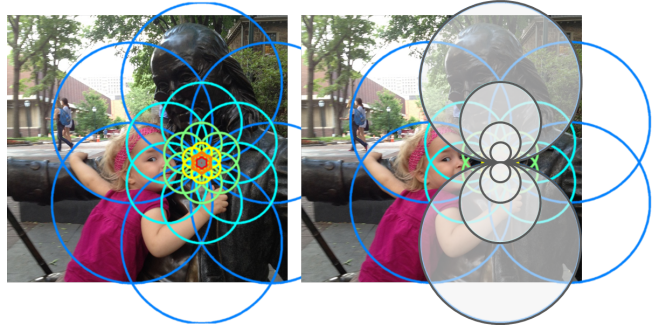


Figure 1. Nested shape descriptors pool scaled and oriented gradients over large geometric structures called *Hawaiian earrings*. (left) *Hawaiian earrings* with k -fold rotational symmetry define a member of the nested shape descriptor family called the *seed-of-life descriptor* (right) Two *Hawaiian earrings* substructures in the seed-of-life descriptor are highlighted in grey.

returns for constructing a descriptor too large. For example, there may be arbitrarily large outliers in the descriptor due to occlusions and geometric variation effects far from the descriptor center. So, an ideal descriptor would be as large as possible, while being robust to occlusions.

In this paper, we introduce nested shape descriptors to address this tradeoff. A nested shape descriptor (NSD) is a family of binary local feature descriptors constructed by pooling oriented and scaled gradients over a large geometric structure called an *Hawaiian earring*. An example of the nested shape descriptor is shown in figure 1. Each descriptor has global support covering the entire image, and the structure of the descriptor exhibits fractal self-similarity in scale. This correlated nested structure enables new a robust distance function called the *nesting distance*. The nesting distance uses order statistics for robustness to outliers while maintaining a descriptor with global support.

Nested shape descriptors make three primary contributions.

- **Global support:** Each NSD exhibits support that covers the entire image, which provides improved selectivity for cases exhibiting the aperture problem without sacrificing localization accuracy.

- **Binary:** NSDs are binary, which enables for compact storage and allows the nesting distance to use a fast Hamming distance, without sacrificing matching performance.
- **Robust local distance function:** The nesting distance is a quadratic local distance function that is robust to corruption of the descriptor due to occlusions, geometric variations or lighting.

In this paper, we provide sufficient conditions for construction of a nested shape descriptor using key concepts of cumulative nested pooling and log spiral normalization. We perform a trade study to determine optimal descriptor parameters for the task of image matching. Finally, we evaluate performance compared to other local feature descriptors on the VGG-Affine image matching benchmark showing measurable performance gains.

2. Related Work

There have been many local feature descriptors proposed in the literature in the past ten years. From oldest to newest, the primary developments have been: SIFT [9], PCA-SIFT [6], Shape context [2], SURF [1], GLOH [11], Sparse localized features (SLF) [14], compressed HoG (cHoG) [4], DAISY [23, 20], BRISK [8], BRIEF [3], ORB [17] and FREAK [15].

The trend in local feature descriptor research has been to show comparable performance to SIFT on the VGG-Affine benchmark [11, 12, 21], with ever faster computation. Work has progressed from PCA-SIFT [6] and SURF [1] which show close performance to SIFT with lower dimensionality and faster preprocessing. Recent work has focused on introducing binary features from local comparison tests [3, 8, 17, 15] which enables fast distance metric based on Hamming distance and faster derivatives [13]. These developments have been driven by the need for faster processing to support mobile deployment.

A taxonomy for comparing and contrasting local feature descriptors can be described in terms of five criteria: preprocessing, support, pooling, normalization and descriptor distance. Preprocessing refers to the filtering performed on the input image, support patterns are the geometric structure used for constructing the descriptor and pooling is the aggregation of filter responses over the support structure. The supplementary material includes a figure which compares dominant local feature descriptors according to these criteria.

Using this taxonomy, the nested shape descriptor is most closely related to DAISY, BRISK and FREAK. NSD has large support and distance robust to occlusions like DAISY, but it does not require an iterative optimization framework determine occlusion masks. NSDs are binary with large support like BRISK/FREAK, however NSD support

is larger and global relative to the image size. Furthermore, unlike BRISK and FREAK, NSD uses scaled and oriented gradients comparisons rather than pixel comparisons for computing the binary representation.

Finally, local distance functions [16] have been explored for metric learning of exemplar distances for the task of object recognition. However, distance functions for local feature descriptors have been limited to Euclidean, Hamming and Mahalanobis distances, where covariance estimation is typically used only for dimensionality reduction [6][11]. In the taxonomy of [16], the nesting distance is per-exemplar (“where”), online (“when”) using order statistics (“how”) without requiring any offline training.

3. Nested Shape Descriptors

In this section, we describe the construction of nested shape descriptors. NSD are constructed by first defining the nested pooling structure (section 3.1), which can be decomposed into a sets of “Hawaiian earring” structures. We provide definitions for this construction and show how the nested shape descriptor is constructed from these pieces (section 3.2). Furthermore, we define the nesting distance (section 3.4), which uses the properties of the nested descriptor to provide robust distance function. Finally, we define a specific member of the nested shape descriptor family called the seed-of-life descriptor (section 3.3), constructed using Hawaiian earrings with k -fold rotational symmetry.

What is the intuition behind the nested descriptor? Figure 2 shows three cases that motivate the use of nesting. The nested descriptor and nesting distance are compared to a generic grid descriptor (e.g. SIFT, but the same argument holds for log-polar grid descriptors) for three common scene variations: occlusions, viewpoint and scale. The red X’s and green checkmarks show where a grid descriptor is corrupted due to the scene variation, which leads to poor matching performance. For these cases, the NSD and nesting distance are able to select the best subset of supports during matching to provide robustness to these scene variations. See the caption in figure 2 for a discussion.

Why the nesting distance? Given a pair of descriptors, the nesting distance computes a weighted sum of the best k coordinate matches. If a coordinate is an outlier (e.g. the worst $n - k$ coordinates, where n is the dimensionality of the descriptor), then any inliers correlated with this outlier are suspect, and are appropriately downweighted. The nesting distance relies on nesting, such that all supports are linked by exactly one point in the center of the descriptor. We discuss further in section 3.4.

3.1. Hawaiian Earrings and Nested Pooling

Nested shape descriptors represent shape using cumulative pooling of oriented gradients within *Hawaiian earrings*. Figure 1 (right) shows an example of the Hawaiian

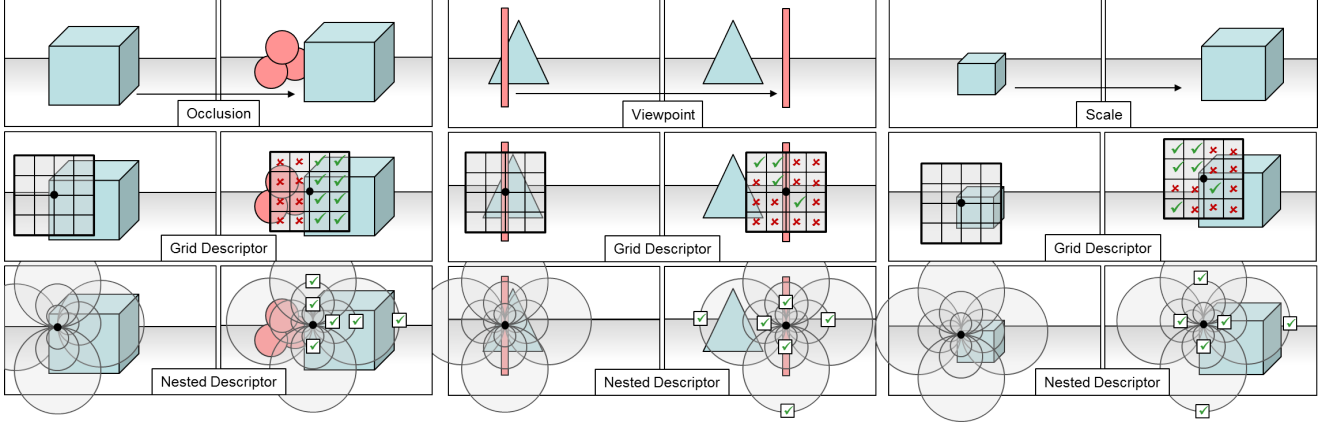


Figure 2. Why nesting? (left) Occlusions corrupt half of a generic grid descriptor covering the occluded region (red X's), while the nesting distance selects the best subset of supports in the nested descriptor that cover only the object (green checkmarks). (middle) Viewpoint changes for long and thin foreground structures introduce errors in grid descriptor matching due to large changes in the background. The nesting distance selects the subset of supports during matching that cover the foreground and are the correct scale to allow for background variation. (right) Scale changes without scale invariant detectors introduce errors in grid descriptor matching due to changes in local support. The nesting distance uses a subset of both large and small scale supports, ignoring intermediate scale supports with corruption.

earring substructure formed by a nested set of circles all intersecting at exactly one point at the center. The Hawaiian earring is a nested structure analogous to Matryoshka or Russian nesting dolls, where each smaller doll fits neatly inside the next larger doll. Hawaiian earrings may be combined into sets such that each earring is called a *lobe*. Each lobe exhibits scale symmetry and all earrings intersect at exactly one point in the center.

In the supplementary material, we formally define the Hawaiian earrings geometric structure. The definitions provide precise construction, however this formality should not obscure the simple intuitive nature of this descriptor. Nested circles of exponentially increasing radius all intersect at exactly one point in the center, and each circle pools oriented gradient responses at a specific scale. Figure 1 shows this common center point in red.

In the remaining sections, we will use the following notation to reference the substructures of Hawaiian earrings. The index $\mathbb{K}_n(u)$ refers to the u^{th} of n Hawaiian earrings, also called the u^{th} lobe. The index $\mathbb{K}(u, v)$ refers to the v^{th} support of the Hawaiian earring $\mathbb{K}_n(u)$. For example, in figure 1 (right), the two lobes highlighted in grey are Hawaiian earrings $\mathbb{K}_6(1)$ and $\mathbb{K}_6(4)$ and the two largest circles are referenced as supports $\mathbb{K}_6(1, 4)$ and $\mathbb{K}_6(4, 4)$.

3.2. Nested Shape Descriptors

A *nested shape descriptor* D at interest point p is defined by nested pooling, logarithmic spiral normalization and binarization of oriented gradients B over a nested support \mathbb{K}_n .

$$d(i, j, k) = \sum_{q \in \mathbb{K}(j, k)} B_{ik}(q) \quad (1)$$

$$\hat{d}(i, j, k) = d(i, j, k) - d(i, j-1, k-1) \quad (2)$$

$$D(i, j, k) = \begin{cases} 1 & \text{if } \hat{d}(i, j, k) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

Equation (1) is *nested pooling*. Let $B_{rs}(q)$ be a bandpass response at pixel q for subband orientation r at octave scale s [19]. The descriptor $d(i, j, k)$ is the pooled response for orientation subband i , lobe j and lobe scale k . Observe that the bandpass octave scale s is equal to the Hawaiian earring support radius k . In other words, support regions with radius 2^k pool orientation subbands over octave scales k . As the support radius increases, the pooling support contains the next smaller support, resulting in nested pooling within a lobe. Figure 3 (left) shows an example of this construction. Equation (1) shows sum-pooling, but we also experiment with max-pooling over a support. Pooling strategies will be defined in the experimental results section.

Equation (2) is *logarithmic spiral normalization*. A logarithmic spiral is a curve that can be written in polar coordinates as $r = ae^{b\theta}$ for arbitrary positive real constants a and b . A nested support set \mathbb{K}_n exhibits a logarithmic spiral when considering neighboring supports. For example, figure 3 (right) shows an example of the logarithmic spiral for \mathbb{K}_6 . Each turn of angle $\theta_i = \frac{2\pi}{6}i$ is a radius of $r_i = 2^i$, which is equivalent to a logarithmic spiral numerically approximated with parameters $a = 1, b = 0.66191$. Figure 3 (right) shows a log-spiral and its reflection $r = ae^{-b\theta}$ forming an elegant flower-like pattern. This pattern encodes the normalization which is a difference of spiral adjacent support, which provides invariance to additive gradient bias.

Equation (3) is *binarization*. A nested shape descriptor can be binarized by computing the sign of (2). This constructs a nested shape descriptor with binary entries.

Figure 5 shows an example of this construction. Nested

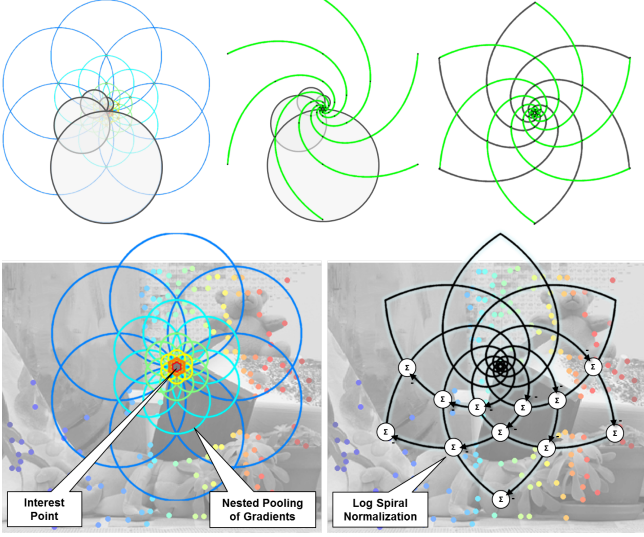


Figure 3. (top) Logarithmic spiral property of the nested shape descriptor provides *normalization* and *binarization*. The log-spiral and its reflection shown in grey form an elegant flower-like structure. (bottom) An NSD is formed at each interest point by (left) nested pooling of scaled and oriented gradients and (right) log-spiral difference and binarization.

pooling is equivalent to pooling of fixed radius over scales of a steerable pyramid [19], which is analogous to a “flattening” of a pyramid representation of scaled and oriented gradients. The final nested shape descriptor D is a binary vector of length $(R \times |\mathbb{K}| \times |K|)$ for R orientation bands over $|\mathbb{K}|$ lobes and $|K|$ supports per lobe. For example, for eight orientation subbands, five nested supports, and six lobes has dimensionality $(8 \times 6 \times 5) = 240$.

3.3. The Seed-of-Life Descriptor

The nested shape descriptors in section 3.2 defines a family of descriptors that share the common properties of nested pooling, log-spiral normalization and binarization. In this section, we define a specific member of this family called the seed-of-life nested shape descriptor or simply the *seed-of-life descriptor*.

The seed-of-life descriptor is a nested shape descriptor such that the nested pooling \mathbb{K}_n is defined using a rotationally symmetric geometric structure called the *seed-of-life*. The seed of life is an ancient geometric symbol formed using Hawaiian earrings with n -fold rotational symmetry. This structure has been discovered as artistic ornamentation in antiquity as far back as the Temple of Osiris in Egypt and Phoenician art from the 9th century BC. It is a central figure in “sacred geometry” where it is a primitive shape used in constructing the “flower of life” and “fruit of life”. An example of the seed-of-life descriptor for \mathbb{K}_6 is shown in figure 1 (left).

The seed-of-life descriptor is perhaps the simplest mem-

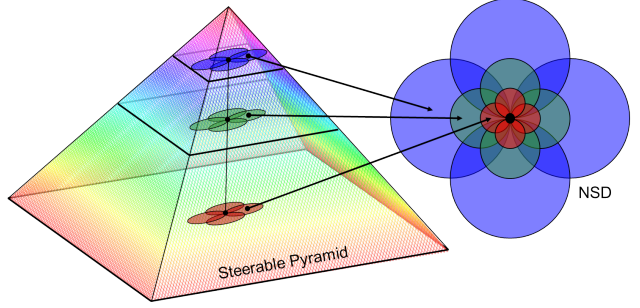


Figure 5. Construction of a nested shape descriptor. An NSD can be considered a “flattening” of the steerable pyramid. Supports of fixed sizes at different levels of the pyramid result in exponentially increasing descriptor supports.

ber of the nested shape descriptor family since it exhibits rotational symmetry where Hawaiian earring lobes are spaced uniformly in angle. This descriptor is used for all experiments in section 4.

3.4. Nesting Distance

The nesting distance is a robust quadratic local distance function [16] unique to NSDs based on *order statistics*. Given two nested descriptors p and q , the nesting distance $d(p, q)$ uses order statistics to partition the supports of two nested descriptors into *inliers* and *outliers* by sorting the squared differences up to a given maximum order k . Then, the nesting distance is equivalent to computing the conditional Gaussian distribution of inliers given outliers.

First, we introduce order statistics. Order statistics are a partial order of a set of variables $\{x_1, x_2, \dots, x_n\}$ such that $x_{(1)} \leq x_{(k)} \leq x_{(n)}$, where the k^{th} order statistic $x_{(k)}$ is equal to the k^{th} smallest value in the set. Common order statistics include the minimum $x_{(1)}$, maximum $x_{(n)}$ and median $x_{(n/2)}$. To simplify notation, we introduce an $n \times n$ binary diagonal selection matrix $S_{(j,k)}$ which encodes a selection of all variables greater than j -order and less than k -order.

$$S_{(j,k)}(i, i) = \begin{cases} 1 & \text{if } x_{(j)} \leq x_i \leq x_{(k)} \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

Observe that using this notation, $S_{(1,k)} + S_{(k+1,n)} = I$, where I is the identity matrix. Furthermore, if all variables are binary, then order statistics take on a simple form

$$x_{(k)} = \begin{cases} 1 & \text{if } \sum_{i=0}^n x_i \geq k, \quad x_i \in \{0, 1\} \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

which is equivalent to a thresholded Hamming distance.

The nesting distance is defined as follows. Let p and q be two nested descriptors of length n . Consider a partition of all squared differences $(p - q)^2$ for a given maximum k -order statistic. Let this partition be represented by selection

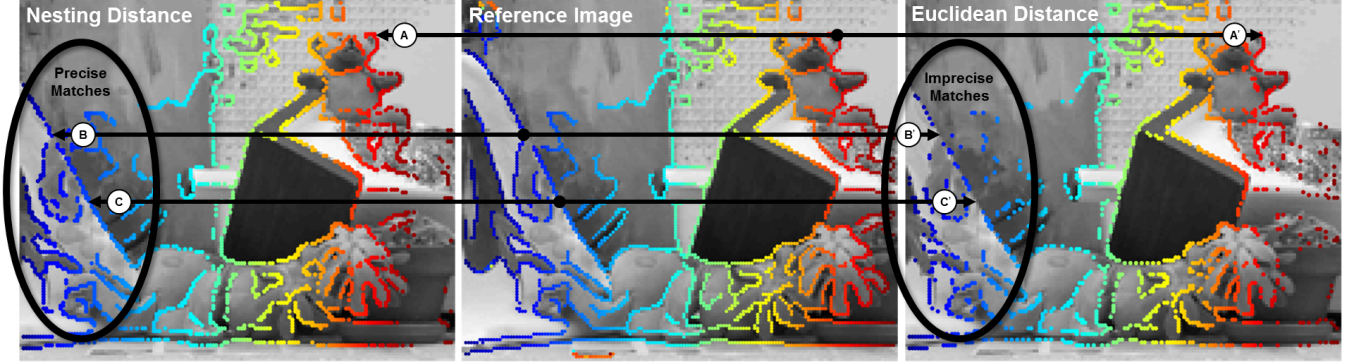


Figure 4. Example stereo image matching using the nesting distance and nested shape descriptors. Colors encode corresponding interest points between the reference image (middle) and the observed image using the nesting distance (left) and Euclidean distance (right). The Euclidean distance is affected by occlusions at the image boundary (left ellipse) resulting in local misalignments, while the nested distance is more robust to these occlusion effects.

matrices of *inliers* $S_{(1,k)}$ and *outliers* $S_{(k+1,n)}$. Then, the nesting distance d is

$$d(p, q, \Lambda, k) = (p - q)^T (I - S_{(k+1,n)}) \Lambda S_{(1,k)} (p - q) \quad (6)$$

where Λ is an optional quadratic weighting matrix. If Λ is diagonal, then this simplifies

$$d(p, q, \Lambda, k) = (p - q)^T \Lambda S_{(1,k)} (p - q) \quad (7)$$

which is simply a sum of k smallest squared differences. Furthermore, if $k = n$ and $\Lambda = I$ then the nesting distance is equivalent to the Euclidean distance.

Lemma 3.1. *If the nesting distance is of the form (6), then it is equivalent to an unnormalized negative log likelihood of a conditional Gaussian distribution for inliers given outliers.*

Proof. We prove this property formally in the supplemental material. Informally, consider a Mahalanobis distance $x^T \Lambda x \propto \mathcal{N}(0, \Lambda)$ as a negative log likelihood (unnormalized) of a Gaussian in canonical form. If a subset of variables are observed (e.g. outliers from order statistics), then well known Gaussian identities can update the conditional likelihood of the remaining variables (e.g. inliers) using the precision matrix Λ for distance weighting. \square

The nesting distance was designed specifically for the structure of the nested shape descriptor. First, recall that there exists exactly one point at the center of the NSD *that is in all supports*. Any subset of supports represents the shape of the center pixel at some orientation and scale. Therefore, this enables the use of order statistics to partition the supports into inliers and outliers, since all supports have one point in common.

The nesting distance cannot be used for descriptors with support constructed on a log-polar or Cartesian grid. Figure 2 (right) shows a simple counterexample. The majority of

the green checkmarks or “good matches” are for supports with no variation on the background far from the center pixel. These matches match the background, are not descriptive for the corner at the center, and *would be the same if we remove the cube altogether*. In contrast, all supports of the NSD include the center pixel due to nesting, so any subset of supports, including large supports, capture the shape of the center pixel.

Figure 4 shows an example of the benefits of the nesting distance for image matching. We extract interest points using an edge based detector, compute nested descriptors at each point, then perform greedy minimum distance assignment from the reference to the observation using either the nesting distance or Euclidean distance. This example shows that the nested distance is more robust to occlusions at the image border than the Euclidean distance.

Finally, The nesting distance has two useful properties that are proven in the supplementary material. First, the nesting distance is non-metric, since it does not satisfy identity or the triangle inequality properties. This property matches perceptual experimentation as it has been long understood that perceptual distance and similarity functions are non-metric [22]. Second, the nesting distance is robust up to corruption of $n - k$ coordinates.

4. Experimental Results

In this section, we provide experimental results for the nested shape descriptor and nesting distance for the task of image matching. First, we perform a trade study using the new experimental protocol of *similarity stereo matching* to determine an optimal set of descriptor parameters for the seed-of-life descriptor. Next, we compare results for the seed-of-life and binary seed-of-life descriptor for the standard VGG-Affine benchmark [12] against SIFT [9] and BRISK [8]. Finally, we show results on a challeng-

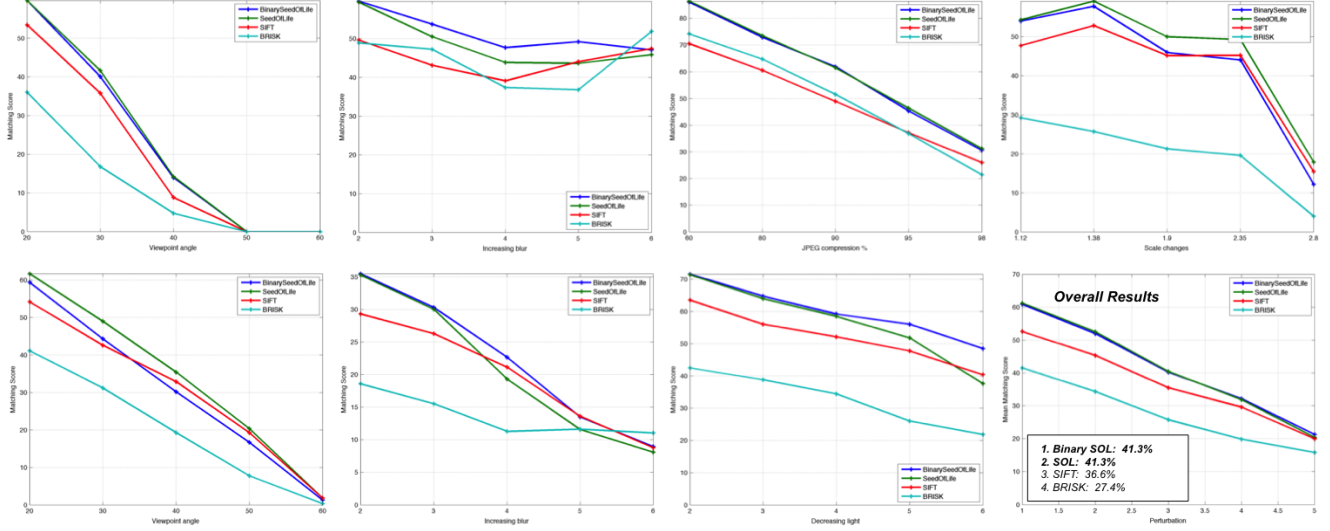


Figure 6. VGG-Affine image matching results. (top) “graf”, “bikes”, “ubc”, “boat”, (bottom) “wall”, “trees”, “leuven” and composite. Both SOL and BSOL outperform SIFT and BRISK, and Binary-SOL is the first binary descriptor to outperform SIFT on this benchmark.

ing application for which traditional local feature descriptors are not applicable. A Matlab toolbox to reproduce these experiments is available at <https://github.com/jebyrne/seedoflife>.

4.1. VGG-Affine

We show comparative performance for local feature descriptor matching on the VGG-Affine benchmark [12]. This dataset includes images of five distortion classes including blur, viewpoint, scale/ rotation, ambient light and JPEG compression. Each distortion class is represented by six images such that the distortion gets progressively worse, and a ground truth homography for performance comparison of local feature descriptors for the task of image matching.

The experimental protocol is as outlined by Mikolajczyk and Schmid [12]. Performance evaluations for local feature descriptors was performed using the matching score criterion (feature match recall). We compare the performance of seed-of-life (SOL) and binary SOL descriptor (section 3.3) to SIFT [9] and BRISK [8]. The seed-of-life is identical to the binary seed-of-life but without the final binarization step of eq. (3). Both SOL and Binary SOL use the Euclidean (and Hamming) distance, as we evaluate the effect of the nesting distance separately in section 4.2. We use a dominant orientation and difference of Gaussians (DoG) scale space feature detector for SIFT and NSD, and the AGAST detector [10] for BRISK. All parameters are defaults provided by the authors, and the parameters for NSD are determined from the trade study in section 4.3, with $k = 0.7n$ for nesting distance. We use the analysis tools and software provided by [12][7][8], and we leave out “bark” for consistency with previous work [8]. The “bark” results are

provided in the supplementary material.

Performance results are shown in figure 6. These results show that either seed-of-life (SOL) and Binary-SOL outperform SIFT and BRISK for all distortion classes. Furthermore, the binary SOL and SOL descriptor perform equally, which shows that the binarization provides a more compact descriptor without sacrificing performance.

4.2. VGG-Affine and Local Distance Functions

Next, we performed a comparison of the nesting distance vs. the Euclidean distance on the VGG-Affine benchmark. This evaluation was proposed to demonstrate the relative benefit of the nesting distance over the Euclidean distance baseline.

Figure 7 shows the results of this study. All distortion classes showed improved performance of the nesting distance over Euclidean. Figure 7 shows matching performance plots for the three distortion classes with the largest benefit, blur (“bikes” and “trees”) and decreasing light (“leuven”). The overall performance shows a 4.1% improvement for the nesting distance.

4.3. Middlebury Stereo and Trade Study

In this section, we perform a trade study to determine an optimal set of parameters for the seed-of-life descriptor. The parameters under study were the number of orientation bands, number of lobes, number of scales, pooling strategy, and binary vs. floating point descriptor elements. We performed a set of studies to understand the effect of these parameters on descriptor performance for the task of image matching.

The experimental protocol for performance evaluation is

detection rate in *similarity stereo matching*. We use six images from the Middlebury stereo dataset [18] (teddy, cones, venus, tsukuba, map and sawtooth). Given a stereo pair (I, J) , ground truth disparity D and similarity transform A , we construct a similarity stereo pair (I, J') such that $J' = A(J)$ by applying the similarity transform A to J . Then, corresponding interest points (p, q) in the similarity stereo pair (I, J') satisfy $p = A^{-1}q + D_p$. Correspondences are the composition of stereo disparity and a similarity transform.

The similarity stereo matching uses the repeatability evaluation protocol of [12] for a range of increasing similarity distortions (scale=0.5-1.5, rotation= $\pm\frac{\pi}{16}$). Random similarities are sampled 10 times for each image at the deformation level and the mean detection rate over all six images for each deformation magnitude is shown.

Figure 8 shows the results of this study. First, we analyzed the effect of the number of orientations. Increasing the number of orientation subbands offers a modest improvement, up to diminishing returns at eight bands. Second, we analyzed the effect of the number of lobes, and found that increasing significantly improves performance up to eight lobes. Third, we analyzed the effect of scales, and found that scale is inversely correlated with deformation. For small deformations, larger scales perform better, but for larger deformations smaller scales perform better. This result summarizes the known tradeoff between descriptor support and matching performance as was discussed in section 1. We selected seven scales. Fourth, we analyzed the pooling strategy and found that sum-pooling (e.g. orientation histograms) had a dramatic improvement over max-pooling for image matching.

The conclusions of this study are a selection of a nominal parameter set. We use eight unsigned orientations, eight lobes, seven scales and sum-pooling. We use these parameters for all experiments in this paper.

4.4. Automated Helicopter Landing

In this section, we describe an application of the nested shape descriptors to the problem of visual landing of a rotary wing platform. Seed-of-life descriptors are used to estimate the position and orientation of a candidate landing zone during approach and landing.

Visual pose estimation for landing is the problem of estimating the 6-DOF position and orientation of a moving landing zone relative to a vehicle with suitable accuracy for safe landing. Given correspondences between an observed image and a known metric markings on the landing zone, we can recover pose using well known techniques of robust homography estimation and decomposition [5].

Performance results are shown in figure 9. We collected four landing approaches of 10Hz 2456x2048 color video and 30Hz 640x512 SWIR video of a manned helicopter ap-

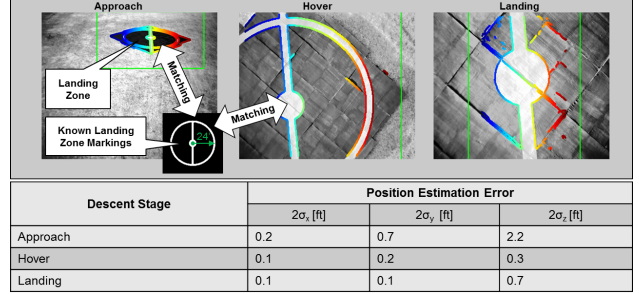


Figure 9. Application of the nested shape descriptors to visual landing zone pose estimation.

proaching a static landing zone during midday. We compared the estimated landing zone position to differential GPS ground truth and results show that the nested shape descriptors achieve 2σ position errors in X, Y and Z of less than 1ft during the descent and landing.

5. Conclusions

In this paper, we introduced the nested shape descriptor family and the associated nesting distance, and showed performance of the seed-of-life descriptor for the task of image matching. Results show that this is the *first binary descriptor* to outperform SIFT on the standard VGG-Affine benchmark. Furthermore, the NSD binary descriptor significantly outperforms BRISK, a state-of-the-art binary descriptor. Future work includes exploring other members of the NSD family such as the flower-of-life or fruit-of-life for improved performance.

Acknowledgments. This material is based upon work supported in part by DARPA under Contract No. W31P4Q-09-C-0051 with Kitware Inc. and NAVAIR contract No. N68335-12-C-0069 with Scientific Systems Company, Inc.

Disclaimers. Approved by DARPA for public release; distribution unlimited. NAVAIR Public Release 2013-389. Distribution Statement A - Approved for public release; distribution is unlimited. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressly or implied, of DARPA, NAVAIR or the U.S. Government.

References

- [1] H. Bay, A. Ess, T. Tuytelaars, and L. Gool. SURF: Speeded up robust features. *Computer Vision and Image Understanding*, 10:346359, 2008.
- [2] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *PAMI*, 24(24):509–521, April 2002.

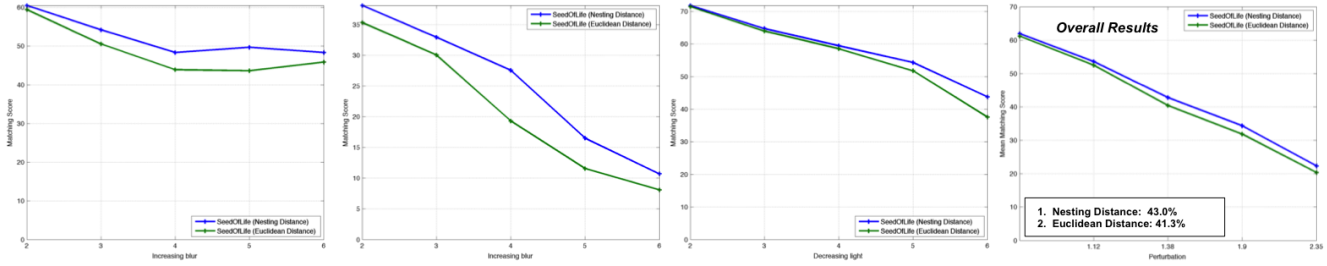


Figure 7. Evaluation of the nesting distance on VGG-Affine dataset. See text for a discussion.

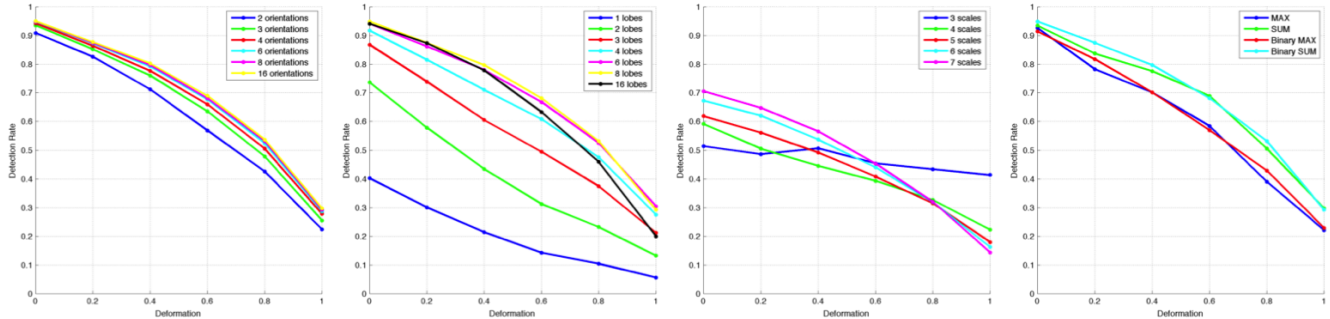


Figure 8. Trade study for the seed-of-life descriptor. (left-right) parameter analysis of orientations, lobes, scales and pooling.

- [3] M. Calonder, V. Lepetit, C. Strecha, and P. Fua. BRIEF: Binary robust independent elementary features. In *ECCV*, 2010.
- [4] V. Chandrasekhar, G. Takacs, D. Chen, S. Tsai, R. Grzeszczuk, and B. Girod. CHoG: Compressed histogram of gradients - a low bit rate feature descriptor. In *CVPR*, 2009.
- [5] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.
- [6] Y. Ke and R. Sukthankar. Pca-sift: A more distinctive representation for local image descriptors. In *CVPR*, 2004.
- [7] K. Lenc, V. Gulshan, and A. Vedaldi. <http://www.vlfeat.org/benchmarks/>, 2012.
- [8] S. Leutenegger, M. Chli, and R. Siegwart. Brisk: Binary robust invariant scalable keypoints. In *ICCV*, 2011.
- [9] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [10] E. Mair, G. Hager, D. Burschka, M. Suppa, and G. Hirzinger. Adaptive and generic corner detection based on the accelerated segment test. In *ECCV*, 2010.
- [11] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *PAMI*, 27(10):1615–1630, 2004.
- [12] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Gool. A comparison of affine region detectors. *IJCV*, 65(1):43–72, 2005.
- [13] M. Muja and D. Lowe. Fast matching of binary features. In *Ninth Conference on Computer and Robot Vision (CRV)*, pages 404 – 410, 2012.
- [14] J. Mutch and D. Lowe. Object class recognition and localization using sparse features with limited receptive fields. *IJCV*, 80(1):45–57, October 2008.
- [15] A. A. R. Ortiz and P. Vanderghelynst. FREAK: Fast retina keypoint. In *CVPR*, 2012.
- [16] D. Ramanan and S. Baker. Local distance functions: A taxonomy, new algorithms, and an evaluation. *IEEE Pattern Analysis and Machine Intelligence (PAMI)*, 33(4):794–806, April 2011.
- [17] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. ORB: An efficient alternative to SIFT or SURF. In *ICCV*, 2011.
- [18] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IJCV*, 47(1-3):7–42, April-June 2002.
- [19] E. Simoncelli and W. Freeman. The steerable pyramid: A flexible architecture for multi-scale derivative computation. In *IEEE Second Int'l Conf on Image Processing*, 1995.
- [20] E. Tola, V. Lepetit, and P. Fua. DAISY: An efficient dense descriptor applied to wide-baseline stereo. *PAMI*, 32(5):815–830, 2010.
- [21] T. Tuytelaars and K. Mikolajczyk. Local invariant feature detectors - survey. *CVG*, 3(1):1–110, 2008.
- [22] A. Tversky. Features of similarity. *Psychological Review*, 84:327–352, 1977.
- [23] S. Winder, G. Hua, and M. Brown. Picking the best DAISY. In *CVPR*, 2009.