

Nested Shape Descriptors - Supplementary Material

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

Descriptor	Preprocessing	Support	Pooling	Normalization	Distance
SIFT	Oriented gradients	Cartesian grid	Histogram	Truncated norm	L2
SURF	Integral image	Cartesian grid	Histogram	Truncated norm	L2
PCA-SIFT	Oriented gradients	Cartesian grid	histogram, PCA	Truncated norm	L2
Shape Context	Edge detection	Log-polar grid	Sum	Global norm	Bipartite matching
GLOH	Oriented gradients	Log-polar grid	Histogram, PCA	PCA	L2
SLF	Laplacian pyramid	Cartesian grid	Max	Global norm	L2
chOG	Oriented gradients	Trees	Histogram	Global norm	L2
DAISY	Convolved orientation maps	Overlapping log-polar	Patch sampling	Support norm	L2
BRIEF	Gaussian filter	Log-polar patch	Gaussian sampled binary comparisons	None	Hamming
BRISK	Gaussian filter	Log-polar patch	Deterministically sampled binary comparisons	None	Hamming
ORB	Gaussian filter	Log-polar patch	Gaussian sampled binary comparisons	None	Hamming
FREAK	Gaussian filter	Log-polar patch	Relatively sampled binary comparisons	None	Hamming
NSD	Steerable pyramid	Nested log-polar	Nested pooling	Log-spiral normalization	Hamming, Nesting distance

Figure 1. Taxonomy and comparison of local feature descriptors.

1. Related Work

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. Figure 1 shows this taxonomy and a comparison of dominant local feature descriptors.

2. Nested Shape Descriptors

2.1. Examples

Figure 2 shows an example of the nesting property. A nested shape descriptor exhibits nesting in two ways, Hawaiian earring nesting and cocentric nesting. An *Hawaiian earring* is a geometric structure formed by the nesting of a set of circles that intersect at exactly one point. The gray opacity in this figure shows that the inner circles are fully contained within the outer circles. Cocentric nesting is formed by a set of nested circles that have the same center. These nesting concepts will be used to construct the nested shape descriptor in this section.

Figure 3 shows an example of the log-spiral pattern

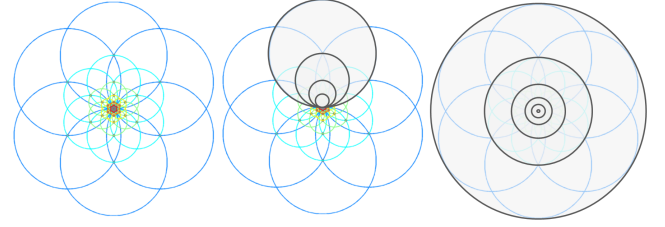


Figure 2. Nesting property of the nested shape descriptor. (left) Seed of life, (middle) Hawaiian earring, (right) Cocentric nesting.

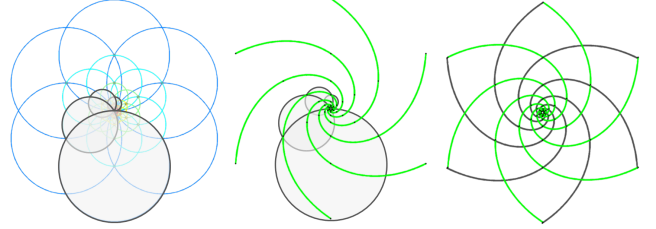


Figure 3. Logarithmic spiral property of the nested shape descriptor provides *normalization* and *binarization*. (right) The log-spiral and its reflection shown in grey form an elegant flower-like structure.

formed by neighboring supports. The sequence of grey circles with centers and radii at left follow the logarithmic spiral shown in green in 3 (middle). Combining this log-spiral with its reflection (right) forms an elegant flower like structure used for normalization and binarization.

Figure 4 shows nested shape descriptors computed for seed-of-life $\mathbb{K}_1 - \mathbb{K}_{10}$. These examples show the rotational symmetry as lobes are added.

2.2. Definitions

In this section, we provide the formal definitions of the Hawaiian earring used to construct the nested shape descriptor.

First, preliminary notation. Let I be an $M \times N$ grayscale image containing pixels $p \in I$ with grayscale value $I(p)$.

Definition 2.1. A support S at c is $S = \{p \mid p \in I, \|p - c\|_2 \leq r\}$

This defines a *support*. Observe that a support is a set

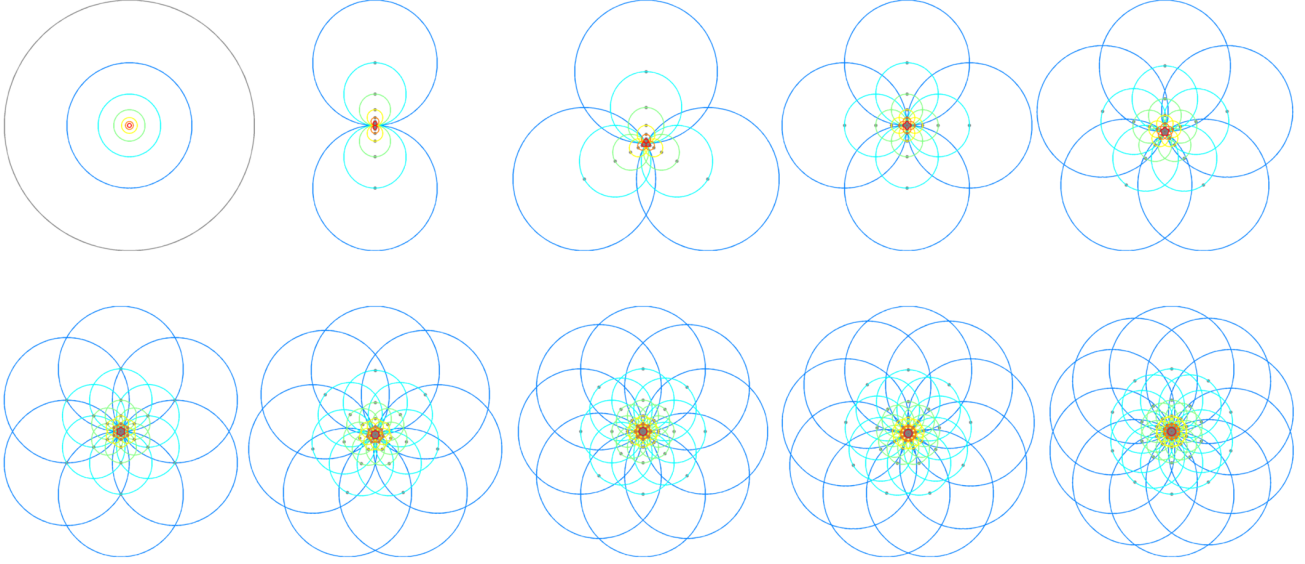


Figure 4. Nested shape descriptors with increasing lobes. (top row) $\mathbb{K}_1 - \mathbb{K}_5$, (bottom row) $\mathbb{K}_6 - \mathbb{K}_{10}$.

of all pixels within a given radius of a center pixel c . For example, each circle in figure 2 is a support.

Definition 2.2. A *nested support set* at p is a set of supports $\mathbb{S}_p = \{S_i \mid r_{i-1} < r_i, p \in S_i, i \leq n\}$ and $S_1 = \{p\}$, $S_n = \{I\}$.

This defines a *nested support set*. A nested support set is an ordered set of supports, such that each support contains the element p and smaller supports are contained within larger supports. Formally, inner support region are strict subsets of all outer support regions, $S_1 \subset S_2 \subset \dots \subset S_n$, radii are totally ordered such that $r_1 \leq r_2 \leq r_n$ and p is contained in each support S_i . The set of grey circles shown in figure 2 (middle) form a nested support set.

The definition of the nested support set implies two useful properties. First, A nested support set is *precise*. It follows from definition (2.2) that the smallest radius $r_1 = 0$ since the innermost support K_1 must contain only p . This definition implies that there exists exactly one point p that is in all supports S_i . This property enables precise pixel level alignment of the nested descriptor for a large support set. Second, a nested support set is *bounded*. The largest support region S_n contains the entire image, which implies that $r_{n-1} < \max(M, N)$ and $r_n \geq \max(M, N)$. This provides a requirement that the largest support must include the entire image to provide global descriptor properties.

Definition 2.3. An *Hawaiian earring* $K(\theta)$ is a nested support set \mathbb{S} such that for each support set $S_i \in \mathbb{S}$, $r_i = 2^i$ and $c_i = (2^{i-1}, \theta)$ in polar coordinates.

This defines a specific case of a nested support set called the *Hawaiian earring*. Each support in the Hawaiian ear-

ring have exponentially increasing radius, the center of each outer circle is on the boundary of the inner circle and all share exactly one common point at the boundary of all circles. The centers of each support are defined in polar coordinates, such that θ is the orientation of the line intersecting all support centers. For example, figure 2 (middle) shows a Hawaiian earring structure in grey, such that the common point is the center of the seed of life structure, and the angle is $K_{\pi/2}$. This structure is fundamental building block of the seed of life and the nested shape descriptor.

Definition 2.4. A *seed of life* \mathbb{K}_n is a set of Hawaiian earrings such that $\mathbb{K}_n = \{K_i(\theta_i) \mid \theta_i = \frac{2\pi i}{n}, \forall i \leq n\}$.

This defines the *seed of life*. This geometric structure is a set of Hawaiian earrings such that each is equally spaced in n polar orientations. Figure 2 (left) shows the seed of life \mathbb{K}_6 for six quantized orientations. The seed of life defines the pooling structure used in the nested shape descriptor and is the primary construction of this section. Figure 4 shows an example of increasing lobes from $\mathbb{K}_1 - \mathbb{K}_{10}$.

2.3. Proofs

In this section, we provide formal proofs for the lemmas referenced in the main body.

Lemma 2.5. If the nesting distance is defined as $d(p, q, \Lambda, k) = (p - q)^T (I - S_{(k+1, n)}) \Lambda S_{(1, k)} (p - q)$, then it is equal to an unnormalized negative log likelihood of a conditional multivariate Gaussian distribution.

Proof. The proof follows by derivation of the nesting distance to the form of an unnormalized conditional Gaussian distribution.

First, preliminary definitions. A joint Gaussian distribution parameterized in canonical form is given by

$$p(x) = \mathcal{N}^{-1}(h, \Lambda) \quad (1)$$

$$f(x) = \frac{1}{2}x^T \Lambda x - h^T x \quad (2)$$

for information vector h and precision matrix Λ . The canonical form $\mathcal{N}^{-1}(h, \Lambda)$ is equivalent to the moment form $\mathcal{N}(\mu, \Sigma)$ using the identities $h = \Sigma^{-1}\mu$ and $\Lambda = \Sigma^{-1}$. The quadratic form (2) follows from the negative log likelihood of the joint density (1), and dropping the constant term.

Let variables x be partitioned into $x = [x_1 \ x_2]$ such that the Gaussian parameters can be partitioned

$$h = [h_1 \ h_2], \quad \Lambda = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix} \quad (3)$$

The conditional distribution $p(x_1|x_2)$ can be derived from the joint distribution $p(x_1, x_2)$ using well known identities.

$$\tilde{h} = h_1 - \Lambda_{12}x_2 \quad (4)$$

$$\tilde{\Lambda} = \Lambda_{11} \quad (5)$$

where $p(x_1|x_2) = \mathcal{N}^{-1}(\tilde{h}, \tilde{\Lambda})$ is the conditional likelihood of remaining variables x_1 given the observation x_2 [3].

Next, we derive a quadratic function g as the conditional likelihood of remaining variables given an observation. To simplify notation, define a selection matrix S that is a binary diagonal matrix that encodes the partition of variables, where $z_1 = S_1x$, $z_2 = S_2x$. With this notation, observe that $x = z_1 + z_2$, and $S_1 + S_2 = I$.

$$g(x) \propto -\log(p(x_1|x_2)) \quad (6)$$

$$g(x) = x_1^T \tilde{\Lambda} x_1 - 2\tilde{h}^T x_1 \quad (7)$$

$$= x^T S_1 \Lambda S_1 x - 2(S_1 h - S_1 \Lambda S_2 x)^T x \quad (8)$$

$$= x^T S_1 \Lambda S_1 x + 2x^T S_2 \Lambda S_1 x \quad (9)$$

$$= x^T (S_1 \Lambda S_1 + 2S_2 \Lambda S_1) x \quad (10)$$

$$= x^T ((S_1 + S_2) \Lambda S_1) x + x^T S_2 \Lambda S_1 x \quad (11)$$

$$= x^T (I + S_2) \Lambda S_1 x \quad (12)$$

This function g is unnormalized negative log likelihood of the conditional distribution, since it drops the constant normalization term.

Finally, the nesting distance d is

$$d(p, q) = (p - q)^T (I - S_{(k+1, n)}) \Lambda S_{(1, k)} (p - q) \quad (13)$$

Let the partition $z_1 = S_{(1, k)}x$ be the set of inliers and $z_2 = S_{(k+1, n)}x$ be the set of outliers determined from order statistics. Then,

$$d(p, q) = (p - q)^T (I - S_1) \Lambda S_2 (p - q) \quad (14)$$

$$d(p, q) = g(p - q) \quad (15)$$

$$d(p, q) \propto -\log(p(x_1|x_2)) \quad (16)$$

□

Lemma 2.6. *The nesting distance is non-metric.*

Proof. We show that the nesting distance satisfies non-negativity and symmetry, but not identity and triangle inequality. Non-negativity $d(P, Q) \geq 0$ is satisfied since all coordinates $(P_i - Q_i)^2$ in the sum are non-negative and real. Symmetry $d(P, Q) = d(Q, P)$ is satisfied since for all coordinates $(P_i - Q_i)^2 = (Q_i - P_i)^2$. Identity $d(p, q) = 0$ iff $p = q$ is not satisfied which can be shown with a simple counterexample. Let $p = [0 \ 0 \ 0]$ and $q = [0 \ 0 \ 1]$, then $d(p, q, \Lambda = I, k = 2) = 0$ but $p \neq q$. Finally, we show a counterexample for the triangle inequality. Let $P = [0 \ 0 \ 0]$, $Q = [0 \ 0 \ 1]$, $R = [1 \ 1 \ 1]$ then $d(P, R, \Lambda = I, k = 2) = 2$, $d(P, Q, \Lambda = I, k = 2) = 0$ and $d(Q, R, \Lambda = I, k = 2) = 1$. Therefore, $d(P, R) \not\leq d(P, Q) + d(Q, R)$ since $2 \not\leq 0 + 1$. □

Lemma 2.7. *If P corresponds to Q , and $\Lambda = I$ then $d(P, Q, \Lambda, k) = 0$ if and only if $\text{corruption}(Q) < \frac{k}{n}$.*

Proof. In this section, “corruption” can be anything that distorts a descriptor such as occlusion, viewpoint, lighting or scale, introducing errors in squared differences in a coordinate during distance computation. Furthermore, a “correspondence” is a true matching of two descriptors P and Q for a given point in a scene.

Let $c = \text{corruption}(Q)$ be a nonzero modification of cN coordinates of Q , where $n = |Q|$. The proof follows from the definition of the nesting distance in that the sum includes the sum of the smallest k squared differences. The largest $n - k$ squared differences can be arbitrarily large without affecting the distance.

(\leftarrow): If $\text{corruption}(Q) < \frac{k}{n}$, then at least k of the coordinates are uncorrupted. Since $P = Q$, an uncorrupted coordinate i has a squared distance $d(P_i, Q_i) = 0$. The bounds of the sum in the nesting distance are the smallest k squared differences, and since at least k are uncorrupted, and each uncorrupted coordinate has distance zero, the sum $d(p, q, \Lambda, k) = (p - q)^T \Lambda S_{(1, k)} (p - q) = 0$.

(\rightarrow): If $d(P, Q, \Lambda, k) = 0$ then the sum of the smallest k squared differences is zero. Since each squared difference is non-negative, each coordinate of the smallest k squared differences must be zero. Therefore, since corruptions are non-zero modifications, the k coordinates are uncorrupted and $\text{corruption}(Q) < \frac{k}{n}$. □

Lemma 2.8. *If P corresponds to Q and exactly one central pixel q of Q is corrupted, then $\text{corruption}(Q) = 1.0$ and $d(P, Q, \Lambda = I, k) > 0$ for all $k > 0$.*

Proof. Let $c = \text{corruption}(Q)$ be a nonzero modification of cN coordinates of Q , where $n = |Q|$. The central pixel q of the nested shape descriptor Q is the center of the nested support set as defined in (2.2). By construction, the smallest radius of the nested support set $r_1 = 0$ since the innermost

support K_1 must contain only q . This implies that there exists exactly one point q that is contained within all supports. Therefore, if q is corrupted, then every support is corrupted. If every support is corrupted, then $\text{corruption}(Q) = 1.0$, then from lemma 2.8 $d(P, Q) \neq 0$, and from the non-negativity property of lemma 2.6 $d(P, Q) > 0$. \square

3. Experimental Results

3.1. Experimental System

In this section, we describe the experimental system used to construct seed-of-life descriptors. The subbands B for a nested shape descriptor are scaled and oriented gradients derived from a complex steerable pyramid [6]. The complex steerable pyramid includes steerable filters in a quadrature pair whose magnitude and phase response are useful for representing signed orientations for "black to white" vs. "white to black" transitions. A Matlab toolbox for building and decomposing separable complex steerable pyramids is available at <https://github.com/jebyrne/sepspyr>.

Max-pooling is performed by max-filtering and sampling and steerable pyramid. First, all bands and scales of the steerable pyramid are 7×7 max-filtered. Then, for each interest point p , we construct lobes by uniform polar sampling of each band at n -orientations at a radius of 3 from p . This sampling proceeds cumulatively over scales, and if a lobe is outside the image, then the cumulative pooling simply uses the nearest valid response. Observe that a 7×7 max-filter at scale i is equivalent to a max-pooled support of size $7 * 2^i$ which allows supports to grow exponentially in size without an exponentially increasing number of pixels in each lobe. Sum-pooling is performed by 7×7 max filtering, followed by summing over spatial support to construct a histogram rather than sampling.

A nested shape descriptor can be similarity normalized using a similarity invariant local feature detector. Given a dominant orientation r^* from a feature detector, a normalizing similarity transform is applied to the seed-of-life pooling structure \mathbb{K} for each interest point. Then, orientation bands are circularly shifted and linearly interpolated such that $D(i', j, k) = \hat{D}(i, j, k)$ and $i' = (i - r^*) \bmod (|R|)$. An analogous approach is used for scale normalization.

A Matlab toolbox for constructing seed-of-life nested shape descriptors is available at <https://github.com/jebyrne/sepspyr>.

3.2. VGG-Affine

Figure 6 shows example feature matching from the VGG-affine dataset. These examples show matched features using NSD and nesting distance for image 2 and image 4 in a subset of distortion classes.

Figure 5 shows the matching score for the "bark" example. This example is commonly left out of evaluations of the

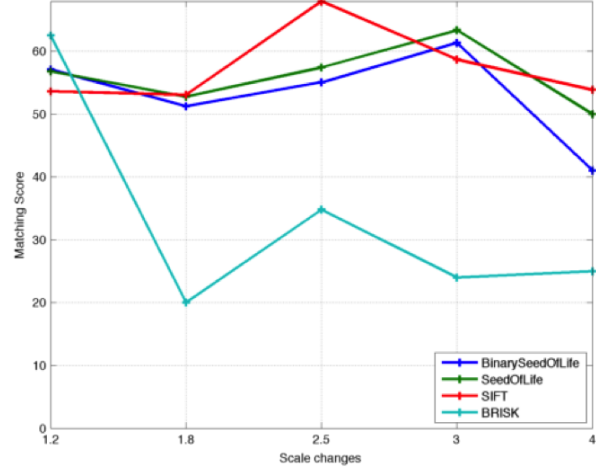


Figure 5. Matching score for "bark" in the VGG-Affine dataset

VGG-Affine dataset as discussed in the main results, since as you can see competing descriptors often perform poorly on this example. However, the results show that the seed-of-life descriptor is competitive with SIFT.

3.3. Automated Helicopter Landing

The zip file contains two videos showing the output of the nested shape descriptors to the problem of pose estimation. These videos show the feature reprojection from linear homography estimation of landing zone markings in a high resolution EO (color) video. The colors encode the matching of interest points defined by edge based detector response from the current image to the first image in the sequence. `perch_29SEP12.mp4` shows the detection (green box) and linear homography reprojection of markings from a point 150 feet aft of the landing zone. `high-over_29SEP12.mp4` shows the detection (green box) and linear homography reprojection from a point 50 feet above the landing zone. These results show that the NSD can robustly detect markings with few corners and scale invariant interest points.

References

- [1] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *PAMI*, 24(24):509–521, April 2002.
- [2] D. Cireşan, U. Meier, and J. Schmidhuber. Multi-column deep neural networks for image classification. In *CVPR*, 2012.
- [3] D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009.
- [4] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, November 1998.
- [5] S. McCann and D. Lowe. Local naive bayes nearest neighbor for image classification. In *CVPR*, 2012.

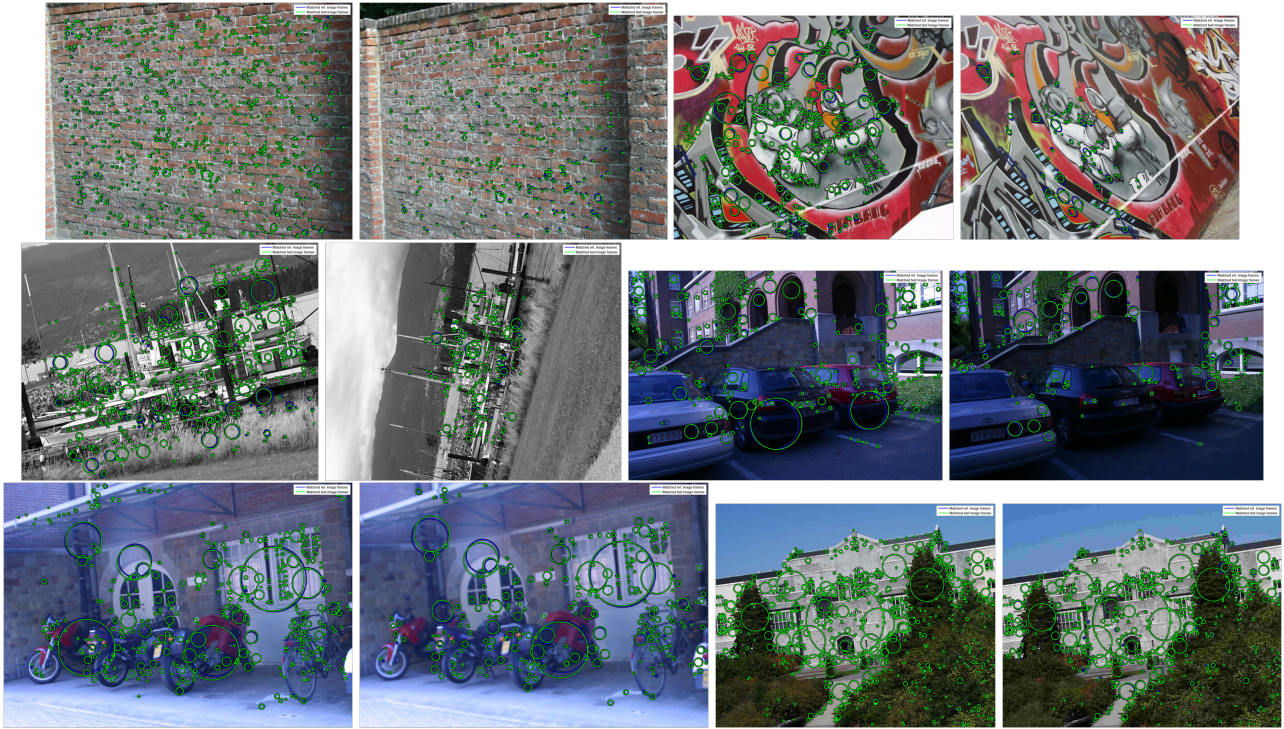


Figure 6. Example matching results from the VGG-Affine dataset.(top-bottom) wall, graf, boat, leuven, bikes, ubc

- [6] 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.