Some title here

Michal Staniaszek

March 4, 2015

Abstract

this is the abstract

Contents

Chapter 1

Introduction

Chapter 2

Background

2.1 Segmentation

Popularised by Comaniciu [15] for use in image segmentation, mean shift was first introduced by Fukunaga [20] in 1975, and rediscovered by Cheng [11] in 1995. The technique finds stationary points in a density estimate of the feature space, for example pixel RGB values, and uses those points to define regions in the space by allocating pixels to them. Pixels which follow the gradient of the density to the same stationary point are part of the same segment. An example can be seen in Figure 2.1.

Random Sample Consensus (RANSAC) is a technique which uses shape models to find ideal models in noisy data. Points in the data set are randomly sampled, and used to construct a shape. For example, in the case of a line, two points are sampled, and define the line. Distances from points in the data set to the model defined by the randomly sampled points are then computed to find points which are inliers to the model. This number is stored, and the process repeated a number of times. At the end of the process, the model with the largest number of inliers is returned [17]. RANSAC can be applied to segmentation tasks by using it to find planes, cylinders, spheres and so on in point clouds. In the case of planes this is particularly useful, as they are usually not part of objects of interest, mostly making up walls, floors or surfaces on which interesting objects rest. By removing the points corresponding to these uninteresting surfaces, segmentation should be made easier. Several extensions to RANSAC have been proposed. Maximum Likelihood Estimation Sample Consensus (MLESAC) chooses a solution that maximises the likelihood of the model instead of just the number of inliers [61]. M-estimator Sample Consensus (MSAC) uses a different cost function to the original implementation, additionally scoring the inliers depending on how well they fit the data [61]. The Progressive Sample Consensus (PROSAC) uses prior information about the likelihood of input data being an inlier or an outlier to limit the sampling pool and greatly reduce computation cost [13].

2.2 Methods for 2D

While we are not directly interested in 2D descriptors, it may be instructive to look at the techniques used in the selection of interest points and the way that descriptors are constructed.

2.2.1 Descriptors

Van Gool [62] gives a description of how to use moment invariants to recognise planar patterns like labels and signs under affine deformations. The moments describe things like the size of the shape and its centre of mass, or statistics like the mean and variance of pixel intensities in the shape. These moments can be combined in such a way that they are invariant to deformations, which is useful to have in a descriptor.

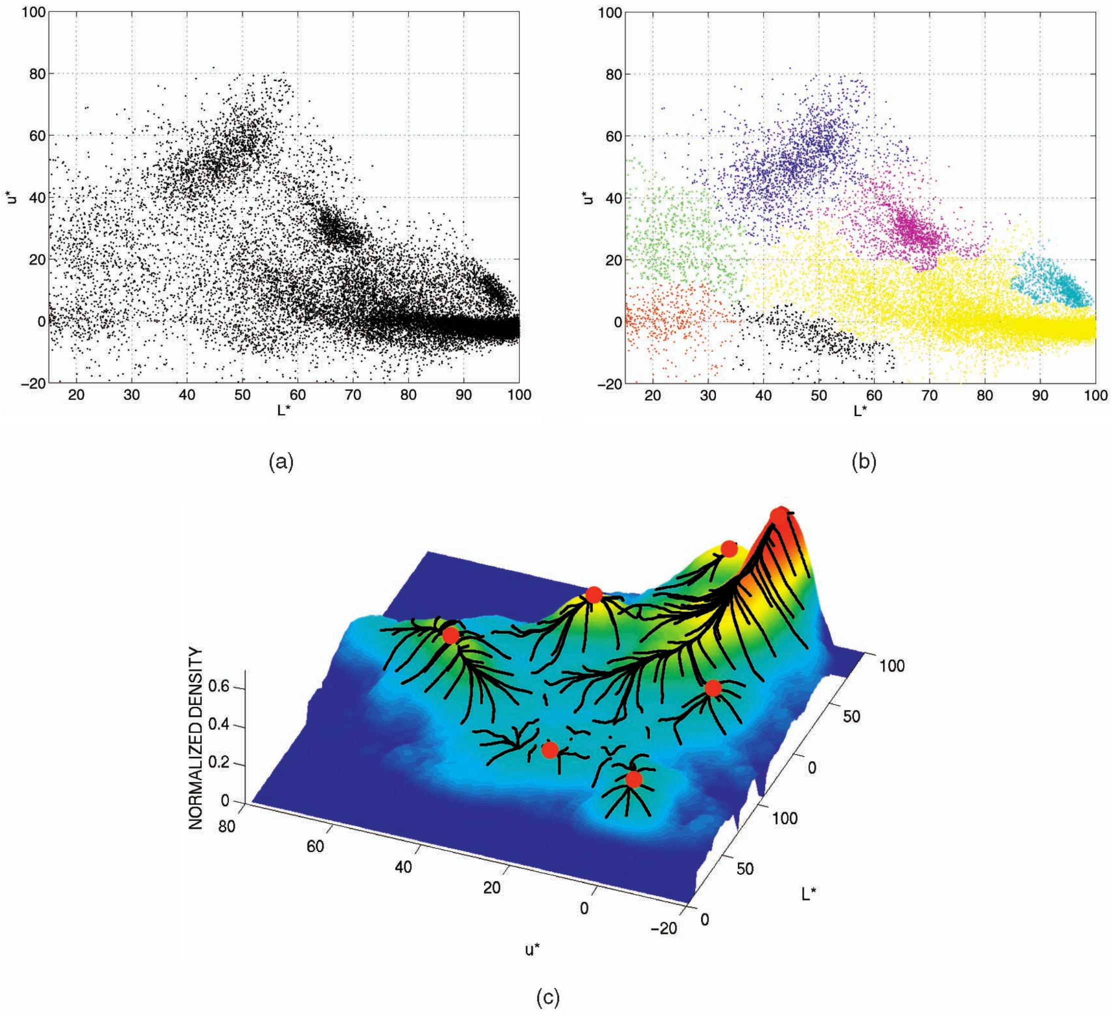


Figure 2.1: Visualisation of mean shift [15]. a) First two components of image pixels in LUV space. b) Decomposition found by running mean shift. c) Trajectories of mean shift over the density estimate.

2.2.2 Combined Interest Point Extraction and Descriptors

The Laplacian of Gaussians was introduced by Lindeberg, and uses derivatives combined with some other techniques to select interest points. [32]. This paper also introduces the concept of automatic scale selection for feature detection, which has played an important part in the field since then. The scale of features can be investigated by blurring an image using a Gaussian kernel — higher standard deviation blurs the image more, resulting in the removal of small scale features.

Even today the Scale Invariant Feature Transform (SIFT) is among the most popular descriptors for 2D images. It is invariant to scale and rotation, and is robust to some variation in affine distortion, viewpoint and illumination, and is distinctive, allowing for correct matching of single features in large databases. There are several stages of computation. Extrama are found in different scales to find points invariant to scale and orientation. Keypoints are selected at the extrema based on their stability. Image gradients at the keypoint are used to define its orientation for future computations. The image gradients are then transformed into a local descriptor vector with length 128 [33].

Mikolajczyk and Schmid [35] introduce the Harris-Laplace detector which is an improvement on SIFT [33] and the Laplacian of Gaussians [32] in the sense that it is able to deal with affine transformations. They do not, however, introduce a new type of descriptor to go with

Figure 2.2: Frames from construction of a spin image [27]. The image plane spins around the oriented point normal and accumulates points.

the point selection.

Speeded-Up Robust Features (SURF) is a more recent descriptor which can be computed and compared much faster than most other descriptors. It makes use of integral images, which replace pixels in an image or image patch with a cumulative sum of the pixel intensities over the rows and columns. This allows for fast computation of pixel intensities in an area of the image. SURF takes some ideas from SIFT, using the spatial distribution of gradients as a descriptor, but integrates over the gradients instead of using individual values, which makes it more robust to noise. The resulting descriptor is a 64 element vector, which means that it is faster to compare than SIFT [3].

2.3 Methods for 3D

2.3.1 Interest Points and Saliency

Sipiran and Bustos extend the popular Harris detector [24] to 3D [52]. Knopp et al. extend the SURF detector to 3D [30].

Shilane and Funkhouser introduce a distinctiveness measure over classes of meshed objects [50].

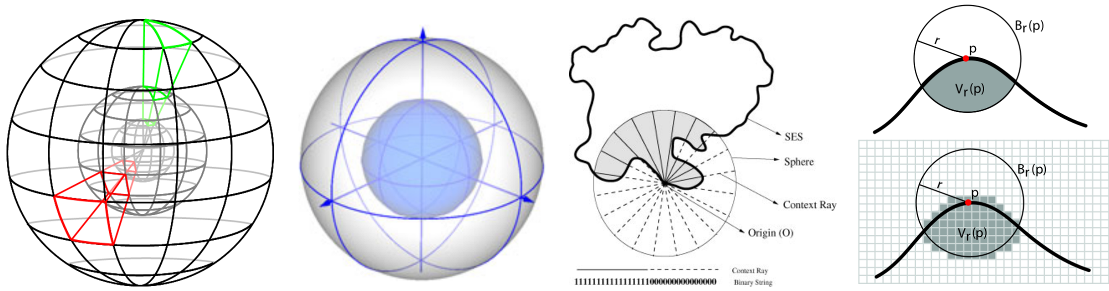
A multi-scale signature defined by the heat diffusion properties of an object called the Heat Kernel Signature (HKS) [57] is used in [43] to retrieve shapes.

2.3.2 Descriptors

One early descriptor which remains popular is the spin image. The descriptor is generated from a mesh model at oriented points with a surface normal. A plane intersecting the normal with a certain width and height is rotated around the normal, forming a cylinder. The plane is separated into bins. The bins accumulate the number of points which pass through a certain bin during the rotation. The resulting 2D image is the descriptor. By varying the width of the plane the region which defines the descriptor can be modified. A small width will give a local descriptor, while a large width will give a descriptor for the whole image [27, 26]. Figure 2.2 shows a visualisation of how the image is generated.

The Ensemble of Shape Functions (ESF) descriptor introduced in [63] by Wohlkinger and Vincze combines the Shape Distribution approach introduced by [42] along with some extensions proposed in [25]. It also makes use of their voxel-based distance measure from [64]. Pairs or triples of points are sampled from segmented partial clouds of objects, and histograms are created by extracting information such as distance, angle, ratios, and whether points are inside or outside (or a mix) of the model. See Figure 2.4.

The Point Feature Histogram (PFH) was introduced by Rusu et al. in [48]. It creates descriptors based on the angles between a point on a surface and k points close to it. The Fast Point Feature Histogram (FPFH) improved the speed of computation, and allowed the use of the descriptor in real time [46]. The Viewpoint Feature Histogram (VFH) extended the FPFH by adding viewpoint information to the histogram by computing statistics of surface normals relative to the viewpoint [47]. It also improved the speed of the FPFH. The clustered version (CVFH) further improved the viewpoint technique by mitigating the effect of missing parts and extending it to facilitate estimation of the rotation of objects [1].



(a) 3DSC [19] (b) SHOT [60] (c) Context Shape [49] (d) Integral Volume [22]

Figure 2.3: Visualisation of spherical descriptors.

Bo et al. develop the kernel descriptor initially created for RGB images for use on depth images and point clouds. The kernels are used to describe size, shape and edge features. Local features are combined to object-level features . Kernel descriptors avoid the need to quantise attributes. Similarity is instead defined by a match kernel [9], which improves recognition accuracy [8].

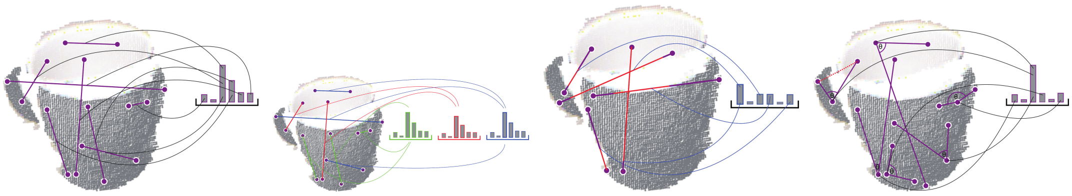
The point pair feature describes the relation between two oriented points on a model. This means that it does not depend so much on the quality and resolution of the model data. The model is described by grouping the point pair features of the model, providing a global distribution of all the features on the model surface [16].

3D Shape Context (3DSC) is an extension of the original Shape Context descriptor for 2D images [5]. A sphere is placed at a point, and its "top" is oriented to match the direction of the normal at the point. Bins are created within the sphere by equally spaced boundaries in the azimuth and elevation, and logarithmically spaced boundaries in the radial dimension (Figure 2.3a). The logarithmic spacing means that shape distortions far from the basis point have less effect on the descriptor. Each bin accumulates a weighted count based on the volume of the bin and local point density [19]. 3DSC does not compute a local reference frame — the vector of the azimuth is chosen randomly, and subdivisions computed from that. This means that a number of descriptors equal to the number of azimuth divisions need to be computed and stored in order to compensate, and the matching process is complicated as a result. The Unique Shape Context (USC) solves this problem by defining a local reference frame and using the directions of that reference frame to subdivide the sphere [59].

The Signature of Histograms of Orientations (SHOT) descriptor improves on 3DSC by taking inspiration from SIFT and making extensive use of histograms. The sphere is split into 32 volumes: 8 azimuth regions, 2 elevation and 2 radial (Figure 2.3b). A local histogram is computed in each of the regions, using the angle between the normal of points and the feature point. The local histograms are then combined to form the final descriptor [60]. The authors also extend the descriptor to include colour (COLORSHOT) [58].

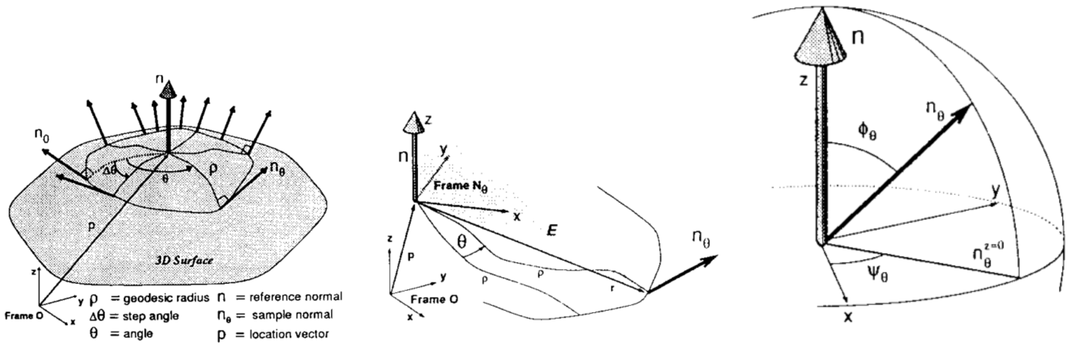
The Rotation Invariant Feature Transform (RIFT) is a generalisation of SIFT. Using intensity values computed at each point from the RGB values, a gradient is computed. Concentric rings are placed around the initial point, and a histogram of the gradient orientations is created for points within each ring. The orientation of the gradient is computed relative to the line from the central point so that the descriptor is rotation invariant. The descriptor is 2D — one dimension is the distance, the other the gradient angle. The distance between two descriptors is measured using the Earth Mover's Distance (EMD), which is a measure of the distance between two probability distributions [31].

Multi-scale descriptors are useful as they can be used to characterise regions of varying size. Cipriano et al. introduce such a descriptor for use on meshes [14]. It captures the statistics of the shape of the neighbourhood of a vertex by fitting a quadratic surface to it. Vertices in the region are weighted based on distance from an initial vertex, and a plane is constructed using



(a) (b) (c) (d)

Figure 2.4: Examples of the measures used to construct the Ensemble of Shape Functions histograms of [63]. a) Distance between points. b) Whether the points are on or off the model, or mixed. c) Ratio of line segments on and off the surface of the model. d) Angle between pairs of lines.



(a) (b) (c)

Figure 2.5: Splash descriptor [56]. a) shows the splash and normals around it. b) and c) show how the additional angles are defined.

a weighted average of the face normals. The parameters of the quadratic are then used to find its principle curvatures, which make up the descriptor.

Work in protein-protein docking also uses 3D descriptors to help with simulations of an otherwise lengthy and complex process. The Surface Histogram is introduced by Gu et al. [23], and uses the local geometry around two points with specific normals on the surface of a protein. A coordinate system is defined by the two points and the line between them, and a rectangular voxel grid is defined around the points. The grid is then marked in locations where the surface crosses the grid, and a 2D image is constructed by squashing the data onto one of the axes. The descriptor is designed to immediately give a potential pose for the docking.

Another example of a shape descriptor from biology is the Context Shape [49]. A sphere is centred on a point, and rays are projected from this point to points evenly distributed on the surface of the sphere (Figure 2.3c). Each of the rays is divided into segments, with a binary value associated with each segment depending on whether the segment is inside or outside the protein. To compare the descriptor, a rotation is applied to match the rays, and a volume of overlap is computed based on matching bits in the rays.

The splash descriptor was introduced by Stein et al. [56]. A point on the surface with a given surface normal (the reference normal) is chosen, and a slice around that with some geodesic radius (distance along the surface) is computed. Points on the circle are selected using some angle step, and the normal at that point is determined. A super splash is when this process is repeated for several different radii. For each normal on the circle, additional angles between it and a coordinate system centred on the reference normal are computed. These angles and the angle around the circle are then mapped into a 3D space, where polygonal approximation is made, connecting each point with a straight line. Some additional computation is done to allow the encoded polygons to act as a hash. Figure 2.5 shows part of the formulation.

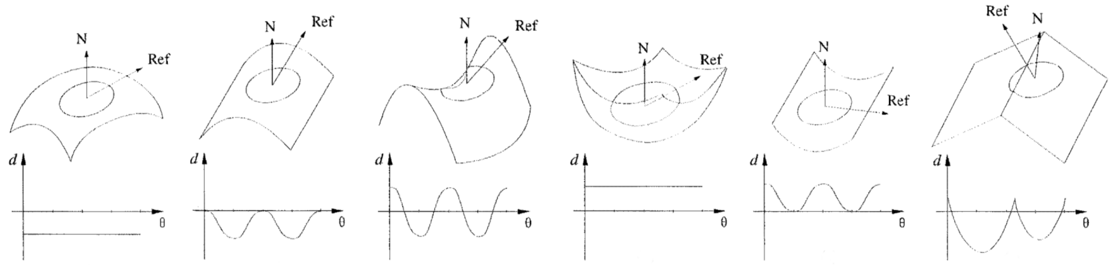


Figure 2.6: Examples of the point signature responses to different surfaces [12]. d is the distance from the reference vector to the space curve defined by the intersection of the surface with a sphere centred at N. Ref rotates about N.

Point Signatures are similar to the splash descriptor in the sense that they both sample points on a circle [12]. This descriptor again selects a reference normal, and has a specific radius. This time, the radius defines a sphere around the point. The intersection of the surface with the sphere is a 3D space curve. The orientation of the curve is defined by fitting a plane to it. The distances between the space curve and the fitted plane at sampled points define the signature of the reference point. These signatures can be compared by lining them up and checking whether the query falls within the tolerance band of previous signatures. Figure 2.6 shows signatures from various surfaces.

2.3.3 Combined Interest Point Extraction and Descriptors

The Normal Aligned Radial Feature (NARF) is an interest point extraction method with a feature descriptor. A score for the image points is determined based on the surface changes at the point, and information about borders. An interest value is computed from this based on the score of the surrounding points. Smoothing is applied, and non-maximum suppression is applied to find the final interest points. To compute the descriptor, rays are projected over the range image from the centre at certain intervals. The intensities of cells lying under the ray are weighted based on their distance from the centre, and a normalised weighted average of the pairwise difference of cells is used to define each element of the descriptor vector, which has a length equal to the number of rays [54]. The method is an improvement on a previous paper by the authors [55]. A problem with this method is that it uses range images directly. Point clouds can be used to generate range images by looking at them from different viewpoints, but this adds complexity to the method.

The integral volume descriptor is interesting as it combines interest point selection and description into one. The descriptor is defined as the volume of the intersection of a sphere centred at a point on the surface of an object with the inside of the object (Figure 2.3d). Interest points are selected by histogramming the descriptor values, identifying bins with a number of points less than a specified values, and selecting points from these bins. To ensure features are properly spaced, points in a certain radius of already selected points cannot be used. By modifying the radius of the sphere used to generate the descriptor, interest points at different scales can be selected [22].

2.4 Storing and Querying Descriptors

There are several techniques for storing and querying descriptors, mostly based on some form of tree. Recently, the k-d tree[6, 18] has been used for efficient approximate matching with either an error bound [2], where there is a bound placed on the error between the true nearest neighbour and the one found, or a time bound [4], where the search is stopped after examining a certain number of leaf nodes. Further improvements on the k-d tree are introduced in [51], where multiple randomised trees are used to optimise the search. A priority search tree algorithm is introduced in [38], which appears to be very effective. This may be the same one as in [37]. The algorithm in the last two papers has been integrated into PCL, which is useful.

A different approach to nearest neighbour search is the balltree, which uses hyperspheres in a hierarchy to enclose points in the space [41]. Unlike the k-d tree, regions on the same level of the tree are allowed to intersect, and no not need to partition the whole space, which gives the balltree its representative power.

The vocabulary tree [40] makes use of techniques from document search to index images. Using k-means clustering, construction stage creates a hierarchical quantisation of the image patch descriptors. In the query phase, descriptors are compared to the cluster centres, and go down the tree until a leaf is reached. The path through the tree is used as a scoring measure to present retrieval results.

Philbin et al. [44] show that flat (single-level) k-means clustering can be scaled to large vocabulary sizes if approximate nearest neighbour methods are used. Early systems for image retrieval used a flat clustering scheme, which could not scale to large vocabularies [53]. The paper also introduces a re-ranking method which uses spatial correspondences, which improves the retrieval quality.

Boiman et al. [10] introduce the Naive Bayes Nearest Neighbour (NBNN) classifier, which seems to be quite relevant. It uses nearest neighbour distances in the space of descriptors instead of images, computing "image-to-class" distances without quantising the descriptors. In general, quantisation allows for dimensionality reduction, at the expense of the discriminative power of descriptors. NBNN "can exploit the discriminative power of both (few) high and (many) low informative descriptors". The problem here is that the classes must be known beforehand, and in our case we do not have that information. The local NBNN [34] does not do the search based on classes. Instead, all the descriptors are merged into a k-d tree on which approximate k-NN is run to find descriptors in the local region of a query descriptor. A distance to classes not present in the k-NN region is approximated by the distance to the $k + 1$th neighbour.

Funkhouser and Shilane present a method for querying a database of 3D objects represented by local shape features [21]. Partial matches (correspondences) are stored in a priority queue sorted by geometric deformation and the feature similarity. This means that only objects in the database with a high probability of being a match need to be processed.

Some work has been done on optimising the retrieval of relevant images by learning from user input [45]. When retrieved images are presented, the user ranks them in terms of relevance, and this rank is then used to improve the relevance of future searches.

2.5 Other

Here are some papers which are interesting and might yield some interesting insights or techniques, but do not directly relate to our problem for whatever reason.

The convolutional k-means descriptor introduced in [7] and BRAND descriptor [39] work on RGB-D images, not directly on point clouds. [29] makes use of the visibility of objects and point pair features to improve recognition performance. Kim et al. use the RGB image and corresponding point cloud to improve the hypotheses for segmentation [28]. [36] uses a growing neural gas to characterise the shape of parts of objects.

Bibliography

[1] Aitor Aldoma et al. “CAD-model recognition and 6DOF pose estimation using 3D cues”. In: *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*. IEEE. 2011, pp. 585–592.

[2] Sunil Arya et al. “An optimal algorithm for approximate nearest neighbor searching fixed dimensions”. In: *Journal of the ACM (JACM)* 45.6 (1998), pp. 891–923.

[3] Herbert Bay et al. “Speeded-up robust features (SURF)”. In: *Computer vision and image understanding* 110.3 (2008), pp. 346–359.

[4] Jeffrey S Beis and David G Lowe. “Shape indexing using approximate nearest-neighbour search in high-dimensional spaces”. In: *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*. IEEE. 1997, pp. 1000–1006.

[5] Serge Belongie, Jitendra Malik, and Jan Puzicha. “Shape matching and object recognition using shape contexts”. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24.4 (2002), pp. 509–522.

[6] Jon Louis Bentley. “Multidimensional binary search trees used for associative searching”. In: *Communications of the ACM* 18.9 (1975), pp. 509–517.

[7] Manuel Blum et al. “A learned feature descriptor for object recognition in rgb-d data”. In: *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE. 2012, pp. 1298–1303.

[8] Liefeng Bo, Xiaofeng Ren, and Dieter Fox. “Depth kernel descriptors for object recognition”. In: *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*. IEEE. 2011, pp. 821–826.

[9] Liefeng Bo, Xiaofeng Ren, and Dieter Fox. “Kernel descriptors for visual recognition”. In: *Advances in Neural Information Processing Systems*. 2010, pp. 244–252.

[10] Oren Boiman, Eli Shechtman, and Michal Irani. “In defense of nearest-neighbor based image classification”. In: *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE. 2008, pp. 1–8.

[11] Yizong Cheng. “Mean shift, mode seeking, and clustering”. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 17.8 (1995), pp. 790–799.

[12] Chin Seng Chua and Ray Jarvis. “Point signatures: A new representation for 3d object recognition”. In: *International Journal of Computer Vision* 25.1 (1997), pp. 63–85.

[13] Ondrej Chum and Jiri Matas. “Matching with PROSAC-progressive sample consensus”. In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. Vol. 1. IEEE. 2005, pp. 220–226.

[14] Gregory Cipriano, George N Phillips, and Michael Gleicher. “Multi-scale surface descriptors”. In: *Visualization and Computer Graphics, IEEE Transactions on* 15.6 (2009), pp. 1201–1208.

[15] Dorin Comaniciu and Peter Meer. “Mean shift: A robust approach toward feature space analysis”. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24.5 (2002), pp. 603–619.

[16] Bertram Drost et al. “Model globally, match locally: Efficient and robust 3D object recognition”. In: *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on.* IEEE. 2010, pp. 998–1005.

[17] Martin A Fischler and Robert C Bolles. “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography”. In: *Communications of the ACM* 24.6 (1981), pp. 381–395.

[18] Jerome H Friedman, Jon Louis Bentley, and Raphael Ari Finkel. “An algorithm for finding best matches in logarithmic expected time”. In: *ACM Transactions on Mathematical Software (TOMS)* 3.3 (1977), pp. 209–226.

[19] Andrea Frome et al. “Recognizing objects in range data using regional point descriptors”. In: *Computer Vision-ECCV 2004.* Springer, 2004, pp. 224–237.

[20] Keinosuke Fukunaga and Larry Hostetler. “The estimation of the gradient of a density function, with applications in pattern recognition”. In: *Information Theory, IEEE Transactions on* 21.1 (1975), pp. 32–40.

[21] Thomas Funkhouser and Philip Shilane. “Partial matching of 3 D shapes with priority-driven search”. In: *ACM International Conference Proceeding Series.* Vol. 256. Citeseer. 2006, pp. 131–142.

[22] Natasha Gelfand et al. “Robust global registration”. In: *Symposium on geometry processing.* Vol. 2. 3. 2005, p. 5.

[23] Shengyin Gu et al. “Surface-histogram: A new shape descriptor for protein-protein docking”. In: *Proteins: Structure, Function, and Bioinformatics* 80.1 (2012), pp. 221–238.

[24] Chris Harris and Mike Stephens. “A combined corner and edge detector.” In: *Alvey vision conference.* Vol. 15. Manchester, UK. 1988, p. 50.

[25] Cheuk Yiu Ip et al. “Using shape distributions to compare solid models”. In: *Proceedings of the seventh ACM symposium on Solid modeling and applications.* ACM. 2002, pp. 273–280.

[26] Andrew E. Johnson and Martial Hebert. “Using spin images for efficient object recognition in cluttered 3D scenes”. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 21.5 (1999), pp. 433–449.

[27] Andrew Edie Johnson. “Spin-images: a representation for 3-D surface matching”. PhD thesis. Citeseer, 1997.

[28] Byung-soo Kim, Shili Xu, and Silvio Savarese. “Accurate localization of 3D objects from RGB-D data using segmentation hypotheses”. In: *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on.* IEEE. 2013, pp. 3182–3189.

[29] Eunyoung Kim and Gerard Medioni. “3D object recognition in range images using visibility context”. In: *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on.* IEEE. 2011, pp. 3800–3807.

[30] Jan Knopp et al. “Hough transform and 3D SURF for robust three dimensional classification”. In: *Computer Vision–ECCV 2010.* Springer, 2010, pp. 589–602.

[31] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. “A sparse texture representation using local affine regions”. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 27.8 (2005), pp. 1265–1278.

[32] Tony Lindeberg. “Feature detection with automatic scale selection”. In: *International journal of computer vision* 30.2 (1998), pp. 79–116.

[33] David G Lowe. “Distinctive image features from scale-invariant keypoints”. In: *International journal of computer vision* 60.2 (2004), pp. 91–110.

[34] Sancho McCann and David G Lowe. "Local naive bayes nearest neighbor for image classification". In: *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on.* IEEE. 2012, pp. 3650–3656.

[35] Krystian Mikolajczyk and Cordelia Schmid. "Scale & affine invariant interest point detectors". In: *International journal of computer vision* 60.1 (2004), pp. 63–86.

[36] C.A. Mueller, K. Pathak, and A. Birk. "Object recognition in RGBD images of cluttered environments using graph-based categorization with unsupervised learning of shape parts". In: *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on.* 2013, pp. 2248–2255.

[37] Marius Muja and David G Lowe. "Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration." In: *VISAPP (1)* 2 (2009).

[38] Marius Muja and David G. Lowe. "Scalable Nearest Neighbor Algorithms for High Dimensional Data". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36.11 (2014), pp. 2227–40.

[39] Erickson R Nascimento et al. "BRAND: A robust appearance and depth descriptor for RGB-D images". In: *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on.* IEEE. 2012, pp. 1720–1726.

[40] David Nister and Henrik Stewenius. "Scalable recognition with a vocabulary tree". In: *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on.* Vol. 2. IEEE. 2006, pp. 2161–2168.

[41] Stephen Malvern Omohundro. *Five balltree construction algorithms.* International Computer Science Institute Berkeley, 1989.

[42] Robert Osada et al. "Shape distributions". In: *ACM Transactions on Graphics (TOG)* 21.4 (2002), pp. 807–832.

[43] Maks Ovsjanikov et al. "Shape Google: a computer vision approach to invariant shape retrieval". In: *Proc. NORDIA* 1.2 (2009).

[44] James Philbin et al. "Object retrieval with large vocabularies and fast spatial matching". In: *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on.* IEEE. 2007, pp. 1–8.

[45] Yong Rui and Thomas Huang. "Optimizing learning in image retrieval". In: *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on.* Vol. 1. IEEE. 2000, pp. 236–243.

[46] Radu Bogdan Rusu, Nico Blodow, and Michael Beetz. "Fast point feature histograms (FPFH) for 3D registration". In: *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on.* IEEE. 2009, pp. 3212–3217.

[47] Radu Bogdan Rusu et al. "Fast 3d recognition and pose using the viewpoint feature histogram". In: *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on.* IEEE. 2010, pp. 2155–2162.

[48] Radu Bogdan Rusu et al. "Persistent point feature histograms for 3D point clouds". In: *Proc 10th Int Conf Intel Autonomous Syst (IAS-10), Baden-Baden, Germany.* 2008, pp. 119–128.

[49] Zujun Shentu et al. "Context shapes: Efficient complementary shape matching for protein–protein docking". In: *Proteins: Structure, Function, and Bioinformatics* 70.3 (2008), pp. 1056–1073.

[50] Philip Shilane and Thomas Funkhouser. "Distinctive regions of 3D surfaces". In: *ACM Transactions on Graphics (TOG)* 26.2 (2007), p. 7.

[51] Chanop Silpa-Anan and Richard Hartley. "Optimised KD-trees for fast image descriptor matching". In: *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on.* IEEE. 2008, pp. 1–8.

[52] Ivan Sipiran and Benjamin Bustos. "Harris 3D: a robust extension of the Harris operator for interest point detection on 3D meshes". In: *The Visual Computer* 27.11 (2011), pp. 963–976.

[53] Josef Sivic and Andrew Zisserman. "Video Google: A text retrieval approach to object matching in videos". In: *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on.* IEEE. 2003, pp. 1470–1477.

[54] Bastian Steder et al. "Point feature extraction on 3D range scans taking into account object boundaries". In: *Robotics and automation (icra), 2011 ieee international conference on.* IEEE. 2011, pp. 2601–2608.

[55] Bastian Steder et al. "Robust on-line model-based object detection from range images". In: *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on.* IEEE. 2009, pp. 4739–4744.

[56] Fridtjof Stein and Gérard Medioni. "Structural indexing: Efficient 3-D object recognition". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14.2 (1992), pp. 125–145.

[57] Jian Sun, Maks Ovsjanikov, and Leonidas Guibas. "A Concise and Provably Informative Multi-Scale Signature Based on Heat Diffusion". In: *Computer graphics forum.* Vol. 28. 5. Wiley Online Library. 2009, pp. 1383–1392.

[58] Federico Tombari, Samuele Salti, and Luigi Di Stefano. "A combined texture-shape descriptor for enhanced 3D feature matching". In: *Image Processing (ICIP), 2011 18th IEEE International Conference on.* IEEE. 2011, pp. 809–812.

[59] Federico Tombari, Samuele Salti, and Luigi Di Stefano. "Unique shape context for 3D data description". In: *Proceedings of the ACM workshop on 3D object retrieval.* ACM. 2010, pp. 57–62.

[60] Federico Tombari, Samuele Salti, and Luigi Di Stefano. "Unique signatures of histograms for local surface description". In: *Computer Vision–ECCV 2010.* Springer, 2010, pp. 356–369.

[61] Philip HS Torr and Andrew Zisserman. "MLESAC: A new robust estimator with application to estimating image geometry". In: *Computer Vision and Image Understanding* 78.1 (2000), pp. 138–156.

[62] Luc Van Gool, Theo Moons, and Dorin Ungureanu. "Affine/photometric invariants for planar intensity patterns". In: *Computer Vision—ECCV'96.* Springer, 1996, pp. 642–651.

[63] Walter Wohlkinger and Markus Vincze. "Ensemble of shape functions for 3d object classification". In: *Robotics and Biomimetics (ROBIO), 2011 IEEE International Conference on.* 2011, pp. 2987–2992.

[64] Walter Wohlkinger and Markus Vincze. "Shape distributions on voxel surfaces for 3D object classification from depth images". In: *Signal and Image Processing Applications (ICSIPA), 2011 IEEE International Conference on.* IEEE. 2011, pp. 115–120.