Probabilistic Object Identification through On-demand Partial Views

Susana Brandão[1,2], Manuela Veloso[3] and João P. Costeira[1]
[1] *Instituto Superior Técnico, Universidade de Lisboa, Av Rovisco Pais, Lisboa, Portugal*
[2] *Electrical and Computer Engineering Department, Carnegie Mellon University, Pittsburgh, U.S.A.*
[3] *Computer Science Department, Carnegie Mellon University, Pittsburgh, U.S.A.*

Keywords: 3D Partial View Representation, Robotic Vision.

Abstract: The current paper addresses the problem of object identification from multiple 3D partial views, collected from different view angles with the objective of disambiguating between similar objects. We assume a mobile robot equipped with a depth sensor that autonomously grasps an object from different positions, with no previous known pattern. The challenge is to efficiently combine the set of observations into a single classification. We approach the problem with a sequential importance resampling filter that allows to combine the sequence of observations and that, by its sampling nature, allows to handle the large number of possible partial views. In this context, we introduce innovations at the level of the partial view representation and at the formulation of the classification problem. We provide a qualitative comparison to support our representation and illustrate the identification process with a case study.

1 INTRODUCTION

We envision mobile robots capable of autonomously recognizing objects in their environments. We assume such mobile robots are equipped with a RGB+D camera, e.g., the Kinect sensor. Such a camera provides only partial views of an object, namely the visible surface of the object. Our goal is to show that a mobile robot can reliably estimate an object class by gathering contiguous partial observations, even when the object is very similar to others. Partial views are collected on-demand by the robot until riching a high confidence on the classification.

We acknowledge that the RGB+D images are inherently noisy and assume that neither the number of observations nor the view angles are known a-priori. However, we do assume that the robot has access to its own motion through its odometry. The proposed identification algorithm is then able to handle arbitrary sequences of noisy observations, constrained only to known changes in the orientation.

We contribute a multiple-hypothesis probabilistic estimation algorithm that updates the robot belief in the object class through noisy observations and own odometry. We start by representing an object, o, as an organized set of partial views by associating each object partial view to a view angle, $\bar{v}$. Thus each partial view corresponds to a tuple $s = (o, \bar{v})$. We then represent each partial view by a noise robust descriptor, $\bar{z}$. To seamlessly handle the series of observations under odometry constraints, $\bar{u}$, we offline learn probability models, $p(\bar{z}|s = (o, \bar{v}))$, for all object classes and view angles. While operating, the robot uses the simple models as building blocks to compose the probability of a sequence of observations. However, since the robot has access to its odometry and not to the absolute view angle, it also needs to estimate the initial orientation. Ambiguities in the descriptor introduced by similarities between objects difficult the initial orientation estimation. We thus use a multiple-hypothesis approach, where we sample possible orientations that are then compared against observations and updated based on odometry.

The proposed algorithm can be described as:

Estimate observation $\bar{z}_1$: From the sensor 3D data, estimate a descriptor $\bar{z}_1$;

Generate M random initial conditions: From all possible objects and orientations, we hypothesize M initial conditions, $s_1^i = [o_i, \bar{v}_i]_1, i = 1, ..., M$;

Compute the probability of each sample: Estimate the probability that each hypothesis generates the observation $\bar{z}_1$.

For each new time step, j: 1. Estimate the descriptor, $\bar{z}_j$;

2. Update hypothesis, $s_j^i = s_{j-1}^i + (0, \bar{u})_j$;

Brandão S., Veloso M. and Costeira J..
Probabilistic Object Identification through On-demand Partial Views.
DOI: 10.5220/0004855507170722
In *Proceedings of the 9th International Conference on Computer Vision Theory and Applications* (VISAPP-2014), pages 717-722
ISBN: 978-989-758-009-3

3. Update the probability for each hypothesis.

We also introduce innovations at the level of representation, $\bar{z}$. Namely we introduce a partial view representation, Partial View Heat Kernel, (PVHK), which is both (i) informative and (ii) robust to the sensor noise. PVHK is informative because it describes the distance between a point centered at the visible surface and each point in the edge in the partial view, as showed in Figure 1. Furthermore, PVHK is robust to noise because it builds upon concepts of diffusive geometry to represent the distances themselves.

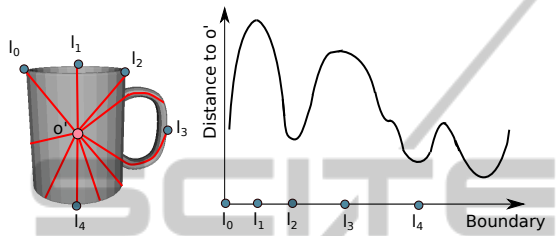


Figure 1: The Partial View Heat Kernel describes a partial view as a function of the distance between a central point, o' and each point in the boundary.

2 RELATED WORK

There is significant research focused on merging information associated with 3D partial views collected from multiple view angles. However it focus on constructing object models. An example is the KinectFusion algorithm (Izadi et al., 2011), which allows the merging of several depth images returned from the Kinect sensor. However, constructing a model does not solve the classification problem.

In this paper, we represent of individual partial views and combine the information at the representation level using a multiple-hypothesis approach. Thus, the related work discussion focus on both classification from multiple instances of the same object and on the representation of individual partial views.

2.1 Multiple-hypothesis on Computer Vision

Multiple-hypothesis approaches have been extensively used for object tracking in 2D color videos, e.g., in (Okuma et al., 2004), or real robots actuating on the environment (Coltin and Veloso, 2011). Furthermore, they have also been extended to include directly object classification as shown in (Okada et al., 2007; Czyz et al., 2007; Hundelshausen and Veloso, 2007). The above approaches separate object position observations from object class observations, in the sense that each corresponds to a set of observations that are represented and handled separately. However, the separation assumes that the objects can be classified independently of the position, which is not the approach we take in the current paper.

2.2 Representations of 3D Partial Views

While the representation of 3D shapes is a very diverse field, we restrict our analysis to representations that describe a complete partial view. Other representations based on local descriptors, such as spin images (Johnson and Hebert, 1999), typically perform worst in noisy scenarios and cannot be used directly in a probabilistic approach.

Approaches to partial views can be divided in three groups. The first describes the partial view based on surface orientations, e.g., as the Viewpoint Feature Histogram (VFH) (Rusu et al., 2010), which represents a partial view by the distribution of surface normals with respect to a central point in the surface.

The second type of representations describes Euclidean geometric properties of the object, e.g., the distances between two points or the mass distribution. The algorithm proposed in (Osada et al., 2002) uses the distribution of Euclidean distances between points on the surface to represent complete objects. By introducing topological information to the distribution, (Ip et al., 2002) made the descriptor more discriminative. Finally, the Ensemble of Shape Functions (ESF) in (Wohlkinger and Vincze, 2011) was introduced by extending the previous approach to partial views.

The added discriminative power resulting from topological information comes at the cost of an increased sensitivity to holes in the object surface resulting from sensor noise. A more robust, but still discriminative, approach relies on the use of diffusive distances (Mahmoudi and Sapiro, 2009), as a noise resilient surrogate to shortest path distances, over the object surface, to represent articulated objects.

Diffusive distances are related with diffusive processes occurring over the object surface, such as heat propagation. Diffusive processes can be interpreted as a sequence of local averaging steps applied to a function representing some quantity, e.g., temperature, defined over some domain. The averaging steps dilute local non-homogeneities in the function and effectively transport the quantity from regions of higher values to regions of lower values. Thus the final distribution of the quantity is generally unaffected by small perturbations caused by noise and topological errors.

The heat propagation was previously used as a basis for building 3D representations. Our proposed descriptor shares with previous work the formalism to

estimate the heat propagation. However, the representation differs significantly as we describe a whole partial view and not a single feature, as the Heat Kernel Signature, (Sun et al., 2009), or a complete object, as the bag of features constructed from Scale Invariant Heat Kernel Signature (SI-HKS), (Bronstein and Kokkinos, 2010). Here, we briefly review the underlying formalism for the heat propagation, however the familiar reader can step to the next section.

2.3 Heat Kernel

Formally, the temperature propagation over a surface, $\mathcal{M}$, of which we have access only to a set of N vertices $V = \{v_1, v_2, ..., v_N\}$ with coordinates $\{\bar{x}_1, \bar{x}_2, ..., \bar{x}_N\}$, is described by eq. 1,

$$L\bar{f}(t) = -\partial_t \bar{f}(t), \tag{1}$$

where $L = \mathbb{R}^{N,N}$ is a discrete Laplace-Beltrami operator and $f_i(t) \in \mathbb{R}$ is the temperature at vertex v_i. We use the *distance* discretization of the Laplace-Beltrami operator, defined by eq. 2:

$$L\bar{f}(t) = (D - W)\,\bar{f}(t), \tag{2}$$

$$W_{i,j} = \begin{cases} 1/\|\bar{x}_i - \bar{x}_j\|^2, & \text{iff} \quad v_j \in \mathcal{N}_i \\ 0, & \text{otherwise} \end{cases}, \tag{3}$$

where D is a diagonal matrix with entries $D_{ii} = \sum_{j=1}^{N} W_{ij}$ and $\mathcal{N}_i$ is the set of vertices that are neighbors to vertex v_i.[1]

The heat kernel is the solution of eq. 1 at vertex v_j when the initial temperature profile, $h(0, \bar{x})$, is a Dirac delta in source vertex v_s. Eq. 4 provides the closed form solution to the heat kernel,

$$k(v_j, v_s, t) = \sum_{i=1}^{N} e^{-\lambda_i t} \phi_{i,j} \phi_{i,s}, \tag{4}$$

where $\phi_{i,j}$ is the value, at vertex v_j, of the eigenvector of L associated with eigenvalue λ_i.

The heat kernel contains information on the complete surface through the eigenvalues and eigenvectors of L. Furthermore, as with other graph Laplacian, $\lambda_1 = 0$ and λ_2 can be seen as the scale of the graph.

(a) t_1 (b) $t_2 > t_1$ (c) $t_3 > t_2$

Figure 2: Heat propagating over an object. Red corresponds to warmer regions and blue to colder ones.

[1]We consider neighborhood relations established from a Delaunay triangulation on the sensor depth image.

3 PARTIAL VIEW HEAT KERNEL

As illustrated in Figure 1, we represent a partial view by the distance between a point in the center of the object and the boundary points. However, we use the heat kernel as a surrogate for the distance for its robustness to noise. In the following we formally define the PVHK and compare it with other descriptors.

3.1 Definition

We define PVHK as the temperature at the boundary measured at some $\Delta t = t_s$ after a source is placed on some vertex, v_s, in the surface. To ensure that PVHK consistently defines a visible surface, we choose v_s as the point closest to the observer, which is also uniquely defined by the tuple $s = (o, \bar{v})$. Additionally, the value of t_s must be large enough to ensure that the heat reaches the boundary but not so long as that all the points are at the same temperature. Since both events depend on the partial view size, and in particular on λ_2, we choose $t_s = \lambda_2^{-1}$.

Thus, given a partial view of an object with a set of vertices V and a set of boundary vertices $B = \{v_{b1}, v_{b2}, ..., v_{bM}\} \subset V$, we compute the temperature at v_{bj} as

$$T(v_{b,j}) = \sum_{i=1}^{\sigma} e^{-\lambda_i/\lambda_2} \phi_{i,bj} \phi_{i,s}, \tag{5}$$

considering only the lowest $\sigma = 30$ eigenvalues, since $e^{-\lambda_i/\lambda_2} \sim 0$ for large i.

Finally, to ensure that all descriptors have the same size independently of the number of vertices, PVHK corresponds to a linear interpolation of the temperature $T(v_{bj})$ with respect to the boundary length. Algorithm 1 summarizes the steps required to estimate the PVHK descriptor.

Data: Set of vertices V, Boundary vertices B, Neighborhoods N, Observer position $\bar{x}_o$.
Result: PVHK descriptor, $\bar{z}$.
Find source position:
$v_s \leftarrow \min_{v \in V} \|\bar{x}(v) - \bar{x}_o\|$;
compute temperature at boundary:
$\bar{T}(v_b) \leftarrow eq.\ 5$;
compute normalized length at each boundary vertex:
$l_B \leftarrow \sum_{j=1}^{M} \|\bar{x}(v_{b,j-1})\|$;
$[\bar{l}]_{i \in \{1,...,M\}} \leftarrow \sum_{j=1}^{i} \|\bar{x}(v_{b,j-1}) - \bar{x}(v_{b,j})\| / l_B$;
interpolate the temperature:
$[\bar{z}]_{k \in \{1,...,K\}} \leftarrow \text{interp1}(k/K, \bar{T}(v_b), \bar{l})$.

Algorithm 1: How to compute PVHK.

The descriptor is stable with respect to perturbations in the object surface, whether from noise or from changes in the sensor position. Thus, descriptors of similar view angles are similar as well. The smoothness of the descriptor variation with respect to the view angle ensures that the complete object can be represented by a finite set of partial views.

3.2 Comparing Representations

We illustrate the potential of PVHK with a qualitative comparison with other partial view representations: the VFH and ESF, from the Point Cloud Library (Rusu and Cousins, 2011) implementation, and the SI-HKS estimated from our own implementation. The analysis focus on the capability for i) distinguish different objects seen from different view angles and ii) for providing a smooth description of partial views.

We thus introduce a partial view dataset constructed from rendering 3D computer models of the rigid objects represented in Figure 3. To simulate realistic spatial and depth resolution as well as noise level, we use a Kinect camera model (Khoshelham and Elberink,). We simulated the camera at 1m from the object and at view angles, $\bar{v} = [\theta, \phi]$, such that ϕ is equal to 45^o and $\theta = 2^o n$, $n = 1, ..., 180$.

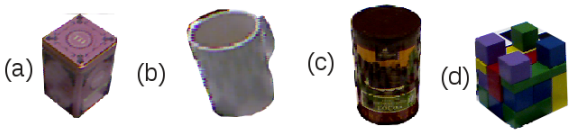


Figure 3: (a) Box; (b) Cup; (c) Cylinder; (d) Castle.

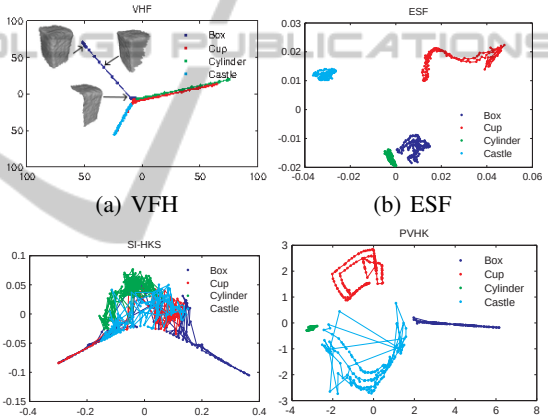
We compare the representations by their 2D isomap projections, (Tenenbaum et al., 2000), represented in Figure 4. Each dot corresponds to the descriptor of a partial view, as illustrated in 4(a), and lines connect those with consecutive view angles. From the projections we see that ESF and PVHK provide robust object representations, since partial views from different objects do not get mixed. However, PVHK is smoother with respect to changes in the view angle. We note that the SI-HKS bag of features approach, while robust for complete objects, is not suitable for describing partial views. Since the heat kernel depends on the complete visible surface, the signature at the same point is affected by changes on the visible surface. The variability resulting from considering the complete set of partial views is not properly reflected by a bag of features approach.

4 SEQUENTIAL IMPORTANCE RESAMPLE FOR OBJECT IDENTIFICATION

Given a noise robust representation, we now address the problem of disambiguating between similar objects, e.g., a glass and a mug.

We start by formulating the problem of object identification as an estimation problem. Namely, our objective is to estimate a sequence of state tuples $s_n = (o, \bar{v}_n)$ from a sequence of observations $\bar{z}_n$ and a set of odometry measurements, $\bar{u}_n$.

In the following, we start by addressing how we model the probability distribution associated with a single partial view. Then we formulate the recognition problem from a set of consecutive partial views as a state estimation problem. Finally we present the main steps to solve the estimation problem using an importance resampling approach.



(a) VFH (b) ESF

(c) SI-HKS (d) PVHK

Figure 4: 2D Isomap projections of different representations of the set of all partial views in the dataset.

4.1 Single Partial View Model

We model descriptors distribution for each partial view as a Gaussian with mean $\mu_{o,\bar{v}}$ and covariance matrix, $\Sigma_{o,\bar{v}}$, $p(\bar{z}|o,\bar{v}) = \mathcal{N}(\mu_{o,\bar{v}}, \Sigma_{o,\bar{v}})$. The distribution reflects the impact of noise on the descriptor and can be learned off-line from empirical data.

4.2 State Estimation

We estimate the object class, o, as the maximum of the a posteriori probability density $p(s_{0:n} = (\tilde{o}_{0:n}, \bar{v}_{0:n})|\bar{z}_{1:n})$, corresponding to the probability of a sequence of states $s_{0:n}$ given a set of observations $\bar{z}_{1:n}$. Since we are interested only in the object class, we marginalize over the view angles $\bar{v}$.

Monte Carlo methods approximate the distribution $p(s_{0:n}|\bar{z}_{1:n})$ by $\sum_{i=1:Np} w_n^i p(s_{o:n}|s_{0:n}^i)$, where $s_{0:n}^i \in \left\{s_{0:n}^0, ..., s_{0:n}^{Np}\right\}$ are a set of support state tuples, i.e., particles, each associated with a weight w_n^i.

In the context of particle filters and a Markovian setting, the object class is estimated as:

$$\hat{o} = \arg\max_{\tilde{o}} \sum_{i=1}^{Np} w_n^i \sum_{j=1}^{Ns} p(s_n^j|s_n^i)\delta_{\tilde{o},o_n^j}, \qquad (6)$$

where $w_n^i \propto p\left(s_n^i|\bar{z}_{1:n}\right)/q\left(s_n^i|\bar{z}_{1:n}\right)$ and $q\left(s_n|\bar{z}_{1:n}\right)$ is the importance sampling distribution from where the particles are sampled at each new time step, n. Finally Ns is the total number of possible states and $\delta_{\tilde{o},o_n^j}$ is a Kronecker delta that ensures that the second sum corresponds to the marginalization over the view angle. The probability $p(s_n^j|s_n^i)$ corresponds to the overlap between the state s_n^j and s_n^i and acts as a kernel between partial views. In practice we estimate it as the confusion matrix between partial views.

4.3 Particles Propagation

The propagation of an initial set of support state tuples, or particles, requires 5 steps:

Step 1: Prediction In this step, we sample particles from the optimal importance density $s_n^i \sim q(s_n|s_{n-1}^i, \bar{z}_n^i) = p(s_n|s_{n-1}^i, z_n)$. We assume a deterministic system dynamics, and thus $q(s_n|s_{n-1}^i, \bar{z}_n^i) = p(s_n|s_n^i)p(s_n^i|s_{n-1}^i)$.

Step 2: Update While the robot moves, the view angle changes as: $\bar{v}_{n+1} = \bar{v}_n + \bar{u}_n$,. We consider that the odometry, $\bar{u}_n = [\delta\theta_n, \delta\phi_n]^T$, is noiseless and so $p(s_{n+1}|s_n, \bar{u}_n) = \delta(\bar{v}_{n+1} - \bar{v}_n - \bar{u}_n)$. Thus, we update the weights as $\tilde{w}_n^i = w_{n-1}^i p\left(\bar{z}_n|s_n^i\right) p\left(s_n^i|s_{n-1}^i\right)$, and $w_n^i = \tilde{w}_{n+1}^i / \sum_{i=1}^{Np} w_{n+1}^i$.

Step 3: Check Resample In this step, we check if the particles have degenerated into a single state. If so, we resample a new set of particles from the current estimation of the probability $p(s_n|z_{1:n})$. The particles degenerate when the number of effective particles, $N_{eff} = N/(1+\sigma(w_n^i))$, is lower than a threshold.

Step 4: Check Restart In this step, the algorithm verifies if at least a particle explains the set of observations. If the non-normalized weights are all smaller than a given threshold δ_{minw}, the algorithm draws a new set of initial particles and restarts the estimation. Assuming that the restart is just a consequence of poor sampling, the algorithm draws new particles for s_0 and updates them using all the past observations $z_{0:n}$.

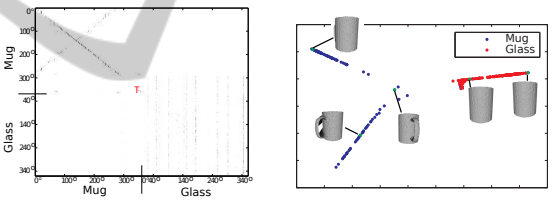
Step 5: Stop Finally, when the variance of the object class probability distribution, $Var(\tilde{o})$, is smaller than a given threshold the estimation process stops.

4.4 Illustrative Example

In this paper we provide insight on both the problem we wish to tackle and the suitability of our approach. For this purpose, we choose a simple example, with just two objects, in detriment of richer and equally successfully examples we could have used.

We thus consider the problem of identifying one of two very similar objects, a mug and a glass, that differ solely on the handle of the former. Since both objects are identical when the handle is not in the field of view, there is a strong ambiguity in the representation. The ambiguity shows both in the confusion matrix and the 2D Isomap projection in Figure 5. Namely, there is a considerable fraction of view angles associated with the mug that are either identified as the glass in the confusion matrix and are completely overlapped in the isomap projection.

Figure 5(b) also highlights the relation between points in the isomap projection and partial views. It shows that the partial views of the mug are separated in three groups: the first is identical to the glass,i.e., presents no handle, the second has a clear handle on the side and the third has a handle at the front.



(a) Confusion matrix (b) Isomap projection

Figure 5: Similarity between a mug and a glass. The red T corresponds to the initial view available for the robot.

We thus hypothesize that a robot acquired the sequence of observations represented on the left column of Figure 6 and odometry measurements equal to $\bar{u}_1 = \bar{u}_2 = [2^o, 0]$. The initial view angle corresponds to the region of ambiguity between the objects. The observations led to three iterations of the algorithm, which we present on the right column of Figure 6 using the isomap projection.

We use 60 particles randomly chosen from a total of 360 states. Each particle is represented by a green square and the current state is represented in a black square. The first observation corresponds to a view angle where the mug and the glass are identical. But on the second and third observations the handle starts to appear on the side and particles jump from the glass branch to the branch with the handle. On the last observation, the large majority of particles is already on the mug branch and the algorithm stops.

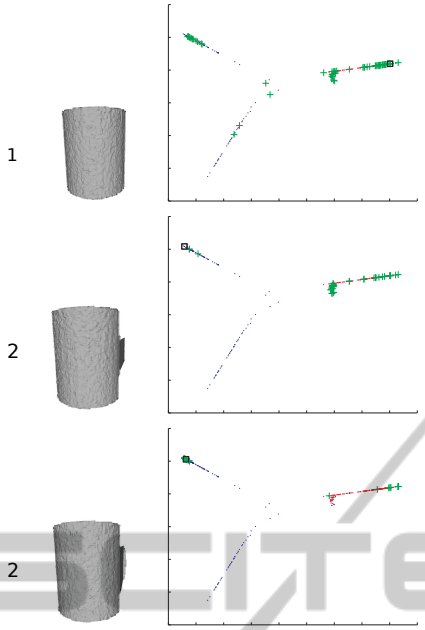


Figure 6: Possible sequence of observations, associated with odometry measurements of $\vec{u}_1 = \vec{u}_2 = [2^o, 0]$ degrees. Green crosses are particles and black square the target state.

5 CONCLUSIONS AND FUTURE WORK

In this work, we presented an algorithm for object identification from multiple partial views. We introduced a sequential importance resampling filter algorithm to combine the set observations. Furthermore, we contribute a descriptor, the Partial View Heat Kernel, to represent the set of observations.

We compared PVHK with other pertinent representations and concluded that PVHK presents several advantages. Namely, we showed that PVHK effectively separates between similar objects and presents smooth variations with respect to changes in the view angle. It is thus suitable in the context of pose estimation since small errors in the descriptor would correspond to small errors in the view angle estimation.

In future steps we propose to test and evaluate the current algorithm with observations captured from a common 3D depth sensor, e.g., the Kinect camera.

ACKNOWLEDGEMENTS

This research was partially sponsored by the Portuguese Foundation for Science and Technology through both the CMU-Portugal and PEst-OE/EEI/LA0009/2013 project, and the National Science Foundation under award number NSF IIS-1012733, and the Project Bewave-ADI. João P. Costeira is partially funded by the EU through "Programa Operacional de Lisboa". The views and conclusions expressed are those of the authors only.

REFERENCES

Bronstein, M. M. and Kokkinos, I. (2010). Scale-invariant heat kernel signatures for non-rigid shape recognition. In *CVPR*.

Coltin, B. and Veloso, M. (2011). Multi-observation sensor resetting localization with ambiguous landmarks. In *AAAI*.

Czyz, J., Ristic, B., and Macq, B. (2007). A particle filter for joint detection and tracking of color objects. *IVC*.

Hundelshausen, F. V. and Veloso, M. (2007). Active monte carlo recognition. In *GCAI*.

Ip, C. Y., Lapadat, D., Sieger, L., and Regli, W. C. (2002). Using shape distributions to compare solid models. SMA.

Izadi, S., Kim, D., Hilliges, O., Molyneaux, D., Newcombe, R., Kohli, P., Shotton, J., Hodges, S., Freeman, D., Davison, A., and Fitzgibbon, A. (2011). Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera. UIST.

Johnson, A. E. and Hebert, M. (1999). Using spin images for efficient object recognition in cluttered 3d scenes. *PAMI*.

Khoshelham, K. and Elberink, S. O. Accuracy and resolution of kinect depth data for indoor mapping applications. *Sensors*.

Mahmoudi, M. and Sapiro, S. (2009). Three-dimensional point cloud recognition via distributions of geometric distances. *Graphical Models*.

Okada, K., Kojima, M., Tokutsu, S., Maki, T., Mori, Y., and Inaba, M. (2007). Multi-cue 3d object recognition in knowledge-based vision-guided humanoid robot system. In *IROS*.

Okuma, K., Taleghani, A., Freitas, N. D., Freitas, O. D., Little, J. J., and Lowe, D. G. (2004). A boosted particle filter: Multitarget detection and tracking. In *ECCV*.

Osada, R., Funkhouser, T., Chazelle, B., and Dobkin, D. (2002). Shape distributions. *ACM Trans. Graph.*

Rusu, R. B., Bradski, G., Thibaux, R., and Hsu, J. (2010). Fast 3D Recognition and Pose Using the Viewpoint Feature Histogram. In *IROS*.

Rusu, R. B. and Cousins, S. (2011). 3D is here: Point Cloud Library (PCL). In *ICRA*.

Sun, J., Ovsjanikov, M., and Guibas, L. (2009). A concise and provably informative multi-scale signature based on heat diffusion. In *SGP*.

Tenenbaum, J., de Silva, V., and Langford, J. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*.

Wohlkinger, W. and Vincze, M. (2011). Ensemble of shape functions for 3d object classification. In *ROBIO*.