

The Partial View Heat Kernel Descriptor for 3D Object Representation

Susana Brandão¹, João P. Costeira² and Manuela Veloso³

Abstract—We introduce the Partial View Heat Kernel (PVHK) descriptor, for the purpose of 3D object representation and recognition from partial views, assumed to be partial object surfaces under self occlusion. PVHK describes partial views in a geometrically meaningful way, i.e., by establishing a unique relation between the shape of the view and the descriptor. PVHK is also stable with respect to sensor noise and therefore adequate for sensors, such as the current active 3D cameras. Furthermore, PVHK takes full advantage of the dual 3D/RGB nature of current sensors and seamlessly incorporates appearance information onto the 3D information. We formally define the PVHK descriptor, discuss related work, provide evidence of the PVHK properties and validate them in three purposefully diverse datasets, and demonstrate its potential for recognition tasks.

I. INTRODUCTION

We address the 3D representation of objects from multiple 3D partial views, where each partial view is the visible surface of the object as seen from a view angle, with no occlusion from other objects. Fig. 1 illustrates a partial view defined in terms of the position of the camera sensor with respect to the object.

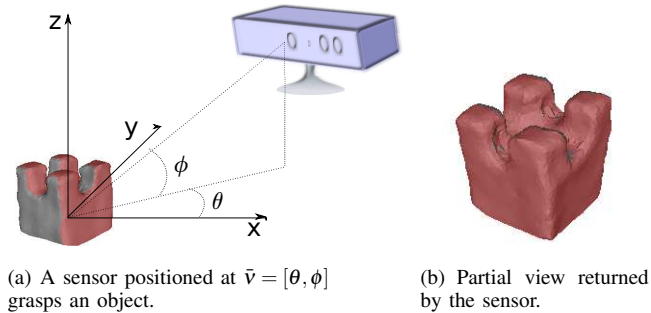


Fig. 1. A partial view is defined by the view angle, \vec{v} , of the sensor with respect to an object centered coordinate system.

Partial views are the returned surfaces of sensors as the object self occludes part of its complete 3D surface. We investigate the 3D representation of the complete object as a set of partial views. We are also interested in common

*Support for this research was provided by the FCT (Portuguese Foundation for Science and Technology) through the Carnegie Mellon Portugal Program under award 22129.1.5004609, by the National Science Foundation award number NSF IIS-1012733, and the Partially funded by the EC through "Programa Operacional de Lisboa"(Project Bewave). The views and conclusions contained in this paper are those of the authors only.

¹S. Brandão (sbrandao@ece.cmu.edu) is with the Electrical and Computer Engineering Department at Carnegie Mellon University, Pittsburgh, USA, and with the Instituto Superior Técnico, Lisbon, Portugal.

²J. P. Costeira (jpc@isr.ist.utl.pt) is with the Electrical and Computer Engineering Department, Instituto Superior Técnico, Lisbon, Portugal.

³M. Veloso (mmv@cs.cmu.edu) is with the Computer Science Department, Carnegie Mellon University, Pittsburgh, USA.

sensor cameras, such as the Kinect sensor, that combine RGB and depth information, but are also noisy. To address these multiple challenges and goals, we introduce a new representation for partial views, the Partial View Heat Kernel (PVHK) descriptor, which is:

- 1) Informative, i.e., a single descriptor robustly describes each partial view;
- 2) Stable, i.e., small perturbations on the surface yield small perturbations on the descriptor;
- 3) Inclusive, i.e., appearance properties, such as texture, can be seamlessly incorporated into the geometry-based descriptor.

The combination of these three characteristics results in a representation especially suitable for partial views captured from noisy RGB+Depth sensors during robot navigation or manipulation where the object surfaces are visible with limited, if any, occlusion.

To ensure that the descriptor is informative, PVHK builds upon a measure related to geodesic-distances to represent distances, over the surface, between a reference point, o' , and the boundary, as illustrated in Fig. 2. The ordered set of all the distances represents surfaces in a unique way, apart from symmetric and isometric transformations.

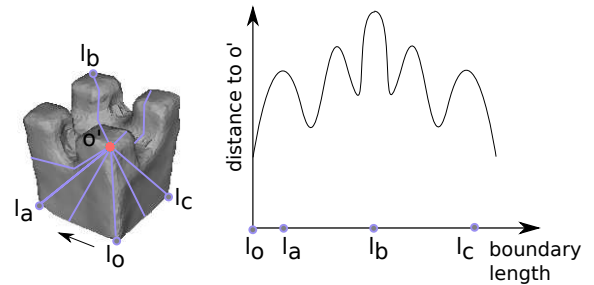


Fig. 2. The Partial View Heat Kernel (PVHK) representation describes a partial view as a function of the distance between a central point, o' , and each point, l_0, \dots, l_c , in the boundary.

To ensure that the descriptor is stable with respect to noise and topological artifacts, e.g., holes or small occlusions, we rely on diffusive geometry concepts to represent *average* distances ([1], [2]). Concretely, we model the averaging process as the propagation along the surface of the heat placed at a source. Hence, PVHK represents a partial view as the temperature at boundary as a result of a heat pulse in the reference point o' .

The stability of the descriptor has a further advantage in that changes in the view angle do not drastically change the surface temperature. Hence, the descriptor varies smoothly with the view angle, which is desirable when representing

objects from partial views. When close view angles generate very different descriptors, the representation of complete objects requires a large number of partial views. By changing slowly, PVHK requires a limited set of partial views.

Finally, to ensure seamless integration of heterogeneous information, such as surface color, PVHK treats different visual properties as different heat propagations velocities. As heterogeneous velocities lead to different temperature profiles on identical 3D shapes, PVHK uniquely represents objects with the same geometry but different color or texture. Furthermore, while color position within the object alters the descriptor, the change is not sensitive to low level details. For example, when applying the descriptor to the representation of people, the change is not sensitive to facial features.

The paper is organized as follows: Section II introduces the datasets used; in Section III discusses related work and reviews the previously introduced Heat Kernel; Section IV introduces our PVHK descriptor, and presents its main characteristics; Section V presents results from two main studied applications; and in Section VI we draw conclusions and address future work directions.

II. DATASETS

In this work, we use three partial view datasets: Dataset-I (Fig. 3), Dataset-II (Fig. 4), and Dataset-III (Fig. 5), which respectively correspond to 3D computer models of rigid objects, to real rigid objects, and to real non-rigid objects.

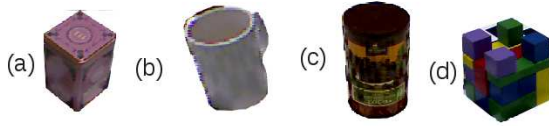


Fig. 3. Dataset-I: (a) box, (b) mug, (c) cylinder and (d) toy castle.

We use Dataset-I to illustrate and compare aspects of both PVHK and other descriptors and therefore require a fine control on the acquisition parameters. We obtain the partial views by rendering the 3D models using spatial and depth resolution similar to that of a Kinect camera. We also add to the depth image a noise proportional to distance, [3], We simulate the camera at a distance of 1m and view angles, $\bar{v} = [\theta, \phi]$, with constant ϕ and $\theta = 2^\circ, 4^\circ, \dots, 360^\circ$.

We use Dataset-II to provide empirical evidence on the accuracy of our representation under real sensor conditions. With a Kinect camera, we collected two sets of partial views, for training and testing respectively, of 13 rigid and of similar size objects. Moreover, the partial views for each object correspond to a known and dense sampling on the observer orientation, $\theta \in [0^\circ, 360^\circ]$.

We use Dataset-III to illustrate the use both color and 3D information and applicability of the descriptor on non-rigid objects. We use partial views of a human in a variety of different positions, also collected using a Kinect camera.

The datasets include objects that share strong similarities in terms of size and shape, e.g., cylinders, cups, and mugs. Furthermore, the objects and partial views have different



Fig. 4. Dataset-II: Real rigid objects.

topologies, to demonstrate the generality of the introduced PVHK descriptor. Using the datasets we intent to illustrate and provide a rich empirical evaluation of our introduced PVHK descriptor.

III. RELATED WORK

We view two ways to represent individual partial views, namely (i) as a set of local features and (ii) as a single holistic feature. We present a brief overview on both alternatives, with emphasis on the holistic features as they provide the background and motivation for PVHK.

A. Local Features

Local features are common to represent partial views, since a small set of features can represent complex objects.



Fig. 5. Dataset-III: Real non-rigid, positions of a human.

For example, Fig. 6 shows the five different features required to represent the box and castle from Dataset-I: three types of corners (P2, P4 and P5), an edge (P3), and a plane (P1).

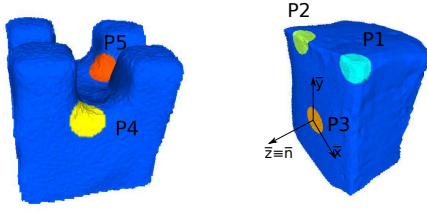


Fig. 6. Example of shapes that can be described using only 5 local features.

Since local representations describe only a small portion of an object, recognition algorithms either solve first a registration problem or combine features into bags of features, similar to bags of words. Consequently descriptors need to be invariant to changes in pose. Several representations achieve this by describing the feature on a tangent space to the object surface at each point, since this space is not only invariant to changes in pose as is easy to reproduce. Examples of such representations are the Fast Point Feature Histogram (FPFH) [4], Signatures of Histograms of Orientations (SHOT) [5], Local Surface Patches (LSP) [6], Spin Images (SI) [7], and Intrinsic Shape Signatures (ISS) [8].

However, methods for estimating the tangent space are sensitive to noise because they rely on normal estimation. As we illustrate in Fig. 7, this negatively reflects on the descriptors. In the figure, we show the variance of different representations as the distance, d , between object and sensor increases, increasing the noise. We estimate the variance by computing the descriptor of the same point over 40 point clouds generated for each value of d . As descriptors have high dimensionality, we represent the variance as ratio between the maximum eigenvalue of the covariant matrix and the mean descriptor. The point used for comparison is $P1$ from Fig. 6 and the descriptors correspond to SHOT, FPFH, and a holistic partial view representation, View Point Histogram, that we include for comparison purposes.

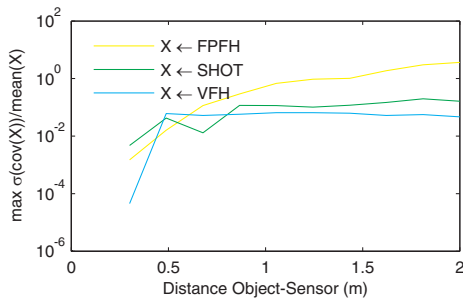


Fig. 7. Noise impact on point like descriptors, SHOT and FPFH, representing the feature $P1$ (plane) from Fig. 6.

B. Holistic Partial View Features

By describing a larger surface, holistic partial view representations are more stable to noise, even when defined on tangent space. E.g., the Viewpoint Feature Histogram (VFH) [9] is an extension of FPFH to the whole partial view but has a lower variance, as we show in Fig. 7.

To altogether avoid tangent space estimation, other representations build upon distances between points on the object surface. E.g., representations for complete objects can be built from the distribution of euclidian distances between points [10]. Extensions to account also for topological information, e.g., [11], are constructed by classifying whether lines connecting pairs of points lay inside the object surface or not. The later was also extended to partial views as Ensemble of Shape Functions, (ESF) [12].

The discriminative power resulting from topological information comes at the cost of an increased sensitivity to holes in the surface due to sensor noise. A more robust approach relies on the use of diffusive distances [1] as a noise resilient surrogate to shortest path distances on object surface.

Diffusive distances are related with diffusive processes occurring over a surface. Diffusive processes can be interpreted as a sequence of local averaging steps applied to a function representing some quantity, e.g., temperature. 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.

Fig. 8 shows two examples of diffusive processes taking place on similar surfaces, different only on account of a hole. In the first example, Fig. 8(a) to Fig. 8(c), the temperature evolves from an initial source to the whole partial view following a concentric pattern, associated with the shortest path between points, until reaching the partial view boundaries. In the second example, Fig. 8(d) to Fig. 8(f), while the hole affects the shortest path between points, it does not strongly effect the temperature. The averaging steps result in that the temperature at a point is defined first by the neighborhood and only implicitly depends on the distance to the source.

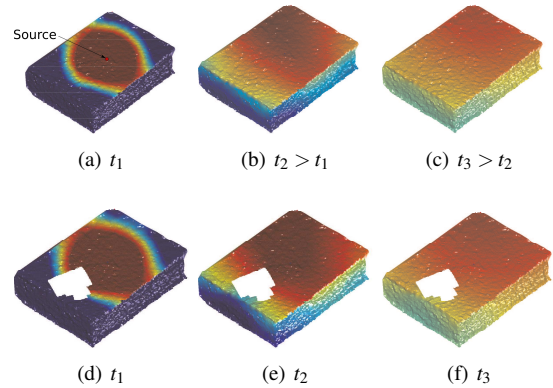


Fig. 8. Example of heat propagation on similar surfaces. Color represents temperature and red regions are warmer than blue. See text for details.

Diffusive processes can describe local features, such as the Heat Kernel Signature (HKS) [2] and the Scale Invariant

Heat Kernel Signature [13]. HKS is a highly robust local descriptor that contains large scale information. HKS represents a point with the temperature evolution after placing, and immediately removing, a heat source on that point. The evolution depends on how fast the temperature propagates to the neighborhood, which in turn depends on the object geometry. While both descriptors, HKS and SI-HKS, perform well on complete 3D shapes, the same point on an object surface may have different descriptors depending on the partial view. Accordingly, matching features across partial views using HKS or SI-HKS is not feasible.

C. Shape and Appearance

Current representations may not extend naturally to include object appearance, i.e., color and/or texture.

Some approaches, e.g., [14], [15], resort to extending ad-hoc the descriptor dimension to include some color/texture descriptor on the extra dimensions. However, the joint descriptors do not effectively associate appearance features with positions in the object.

On the other hand, the photometric heat kernel [16], directly associates appearance to 3D coordinates by changing the space where the object is defined. I.e., each point in the surface lays in a 6D space with physical coordinates plus RGB values. The formalism used for diffusive process extends naturally to this new space, however it takes into account only color gradients and not absolute information, which may hinder recognition.

Given all the above, we see that heat kernel provides a noise resilient representation of distances and has already proved to easily extendable to include color. Thus, we use the heat propagation as a substitute for a set of representative distances over a partial view: the ordered set of distances between boundary points and a single point in the object surface, as we previously illustrated in Fig. 2. By its importance to our proposed representation, we here briefly review the heat kernel and its underlying process. (The familiar reader can step to the next section.)

D. Heat Kernel

Formally, the temperature propagation over a surface defined by a set of vertices $V = \{v_1, v_2, \dots, v_N\}$, with coordinates $\{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N\}$ together with a set of neighborhood relations \mathcal{N} , is described by eq. 1:

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

where $L = \mathbb{R}^{N,N}$ in eq. 2 is a discrete Laplace-Beltrami operator, and $f_i(t) \in \mathbb{R}$ is the temperature at v_i and instant t .

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

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

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

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

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 . The problem has a closed form solution given by eq. 4

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

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

Eq. 4 contains information on the complete surface through the eigenvalues and eigenvectors of L , i.e., even when v_j and v_s are fixed points on the object surface, the descriptors changes if L changes.

Furthermore, as with other graph Laplacians, $\lambda_1 = 0$ and λ_2 can be seen as the scale of the graph.

IV. PARTIAL VIEW HEAT KERNEL

We define PVHK as the temperature at the surface boundary t_s seconds after placing the source on a vertex, v_s .

To ensure that PVHK consistently defines a visible surface, we choose v_s as the point closest to the observer, as we can easily find the point for each view angle. Additionally, the value of t_s must be large enough to ensure that the heat has time to reach the boundary, but also that the equilibrium is not reached, i.e., that not all the points in the surface are at the same temperature. Since both events depend on the partial view size, and in particular on λ_2 , we choose $t_s = \lambda_2^{-1}$.

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

$$T(v_{bj}) = \sum_{i=1}^{30} e^{-\lambda_i/\lambda_2} \phi_{i,bj} \phi_{i,s}, \quad (5)$$

using the lowest 30 eigenvalues, as $e^{-\lambda_i/\lambda_2} \sim 0$ for $i > 30$.

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

Input: Set of vertices V , Boundary vertices B , Neighborhoods \mathcal{N} , Observer position \bar{x}_o .

Output: PVHK descriptor, \bar{z} .

Find source position:

$v_s \leftarrow \min_{v \in V} \|\bar{x}(v) - \bar{x}_o\|;$

compute temperature at boundary:

$\bar{T}(v_b) \leftarrow \text{eq. 5};$

compute normalized length at each boundary vertex:

$l_B \leftarrow \sum_{j=1}^M \|\bar{x}(v_{b,j-1})\|;$

$[\bar{l}]_{i \in \{1, \dots, M\}} \leftarrow \sum_{j=1}^i \|\bar{x}(v_{b,j-1}) - \bar{x}(v_{b,j})\|/l_B;$

interpolate the temperature:

$[\bar{z}]_{k \in \{1, \dots, K\}} \leftarrow \text{interp1}(k/K, \bar{T}(v_b), \bar{l}).$

Algorithm 1: How to compute PVHK for a partial view.

A. Stability and Smoothness

The stability ensures that PVHK changes smoothly with respect to changes in the view angle. Changes in the view angle lead to (i) the source moving over the object, as the closest point to the observer changes; and (ii) the visible surface of the object, including the boundary, changing. As long as there are no severe changes on the object surface, both changes happen smoothly.

The smoothness impact is two fold: (i) it guarantees that the set of descriptors for a given object are contained in a finite and contiguous portion of the descriptors space, which is important for classification with discriminative methods; (ii) it guarantees that descriptors associated with similar view angles are also similar, which reduces the error on applications with view angle estimation.

In Fig. 9 we highlight both characteristics in PVHK on 2D Isomap projections, [17], of four representations applied to Dataset I. Dots corresponds to partial views, as illustrated in 9(a) for the rectangular box, and connected dots are contiguous view angles. The other representations are the VFH and ESF, from the Point Cloud Library [18] implementation; and SI-HKS, from our own implementation. From the projections we see that ESF and PVHK are more effective at separating objects, since partial views from different objects do not get mixed in a 2D projection. However, ESF does not change as smoothly with the view angle as PVHK, notably in the cup and the castle example.

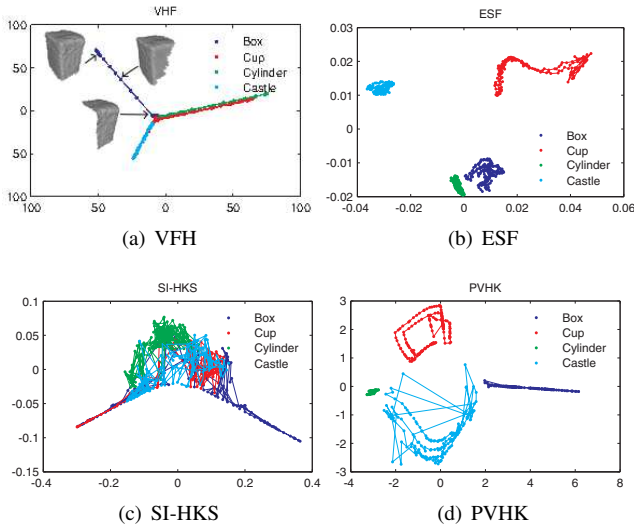


Fig. 9. 2D Isomap projections for four representations applied to all partial views generated from Dataset-I. See the text for details

We also note that SI-HKS bag of features approach for combining local features, while robust at discriminating between complete objects, is not suitable for partial views. As the heat kernel signature of a single point depends on the complete visible surface, there is a greater variability in the descriptor from changes in the view point.

B. Adding Texture

We introduce texture into PVHK representation by slightly modifying the heat equation. The heat equation assumes the whole surface propagates heat at the same rate. By locally using different rates, we generate different the descriptors for objects with the same geometry. Thus, to differentiate objects on both appearance and geometry, we associate appearance with propagation rate. We thus rewrite the heat equation as:

$$Lh(v, v_s, t) = c(v)\partial_t h(v, v_s, t), \quad (6)$$

where $c(v) : V \rightarrow \mathbb{R}$ is any scalar function associated with color, or texture, at vertex v .

The solution to the non-homogeneous problem in Eq. 6 is identical to the solution to the homogeneous problem in Eq. 1, but ϕ and λ are now the solution of the generalized eigenvalue problem $L\bar{\phi}_i = C\bar{\phi}_i\lambda_i$, where C is a diagonal matrix with entries equal to $c(v)$. The solution then becomes $\tilde{h}(v_j, v_s, t) = \sum_{i=1}^M \phi_{i,j} \exp(-\lambda_i t) \phi_{i,s} C_{s,s}$.

Our proposed approach differs from previous efforts to combine color and geometry, in particular from [16]. Notably, we can extend the function $c(v)$ to represent any scalar quantity and not just color. Examples of useful scalars are the color hue value or cluster indices, e.g., after some clustering preprocessing using any other appearance representation.

We illustrate the impact of adding appearance information to the descriptor by considering a person in the same position, wearing the same clothes but with different colors, as in Fig. 10. We assume that our scalar function $c(v)$ corresponds to the color hue value when scaled to the interval $[0.5, 1]$ and present the temperature along the boundary in the graphic on Fig. 10. The four descriptors present a common behavior associated with shape, e.g., the head, point l_1 , introduces the same decrease in the temperature. However, the color modulates the temperature in a very significant way. Notably, the color at the source, which in the example is placed in the blouse, leads to the gap between *Original+Different Skirt* and *Different Blouse + Different Dress*.

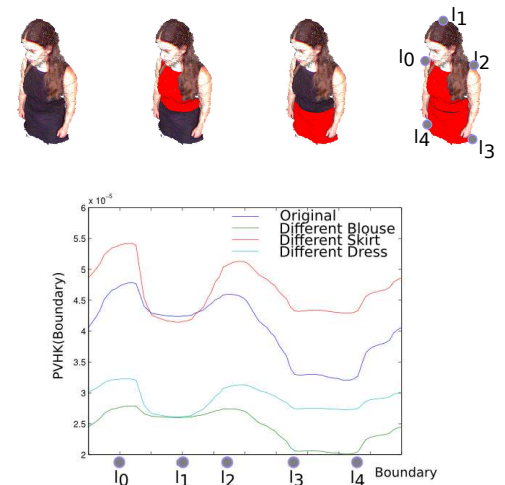


Fig. 10. Color impact on the descriptor. On top, we present the mesh and colors. On the bottom, we present the respective descriptors, i.e., the temperature values along the boundary.

V. APPLICATIONS

We envision mainly two applications for PVHK: (i) recognition of rigid objects, and (ii) tracking of articulated shapes.

A. Recognition

We show the effectiveness of PVHK at object recognition on Dataset-II, without using color. Classification was based on nearest neighbors, using an L2-norm. As PVHK is defined over a closed curve, it is ambiguous over circular shifts. Thus the nearest neighbor search considers all such shifts.

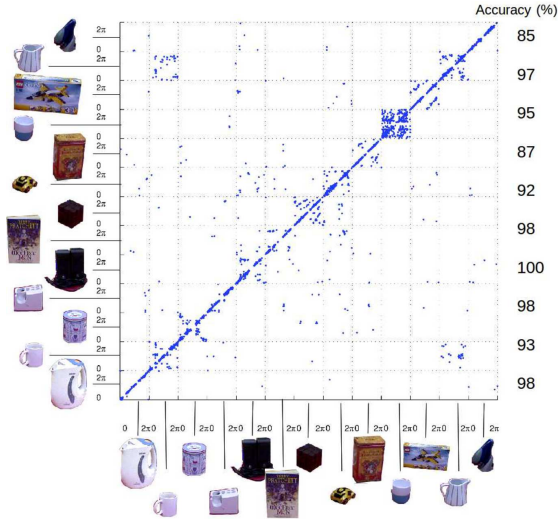


Fig. 11. Confusion matrix for PVHK testing

Fig. 11 presents the a confusion matrix that relates the true view angle of each element on the testing dataset, on the x-axis, to the view angle of the closest descriptor from the training dataset, on the y-axis. The confusion matrix shows that most miss classifications correspond to similar objects, e.g., the cream pitcher and the mug. The matrix shows also the inner category confusion that we expect in objects with strong symmetries, such as the drinking glass. The overall accuracy was 95% and the accuracy for each class is represent in the column to the right of the matrix.

B. Tracking of Articulated Shapes

Deformations on clothes and body due to movement make tracking articulated shapes, such as humans, very challenging. However, the heat kernel is invariant to isometric deformations, which means that PVHK will also be resilient to most deformations and not change with movement.

Using the Dataset-III, we estimate the PVHK descriptor for different poses of a human moving between the six positions in Fig. 5. Fig. 12 presents the 2D Isomap projection and respective labels for the dataset. Specifically, the figure shows the separation of descriptors into two groups associated with the position of the arms with respect to the body.

The results in Fig. 12 imply that, while using our proposed descriptor PVHK, an articulated body can be represented by a reduced number of rigid shapes. Consequently, hint that tracking and recognition, if aided by including color or texture information, can be easily achieved.

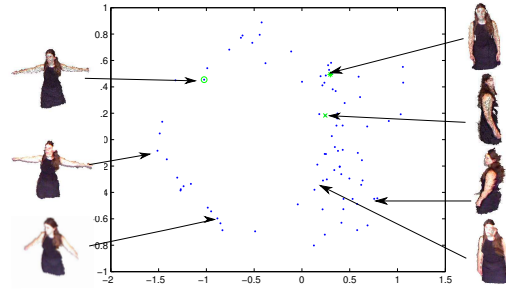


Fig. 12. 2D Isomap projection for a human moving.

VI. CONCLUSIONS

In this work we introduce a partial view representation and demonstrate that it is informative, stable to perturbations and effective at representing both geometry and appearance.

PVHK is especially suitable when there is little occlusion from other objects. We thus foresee a large spectrum of applications ranging from robot manipulation, where there is only the target object present, to robot controlled perception, where the robot can move to avoid occlusions.

Furthermore, the color extension of PVHK is able to represent color distributions over geometry, which opens the door to many other applications. We highlight its adequacy for tracking humans in a contained environment, where, by being insensitive to low level details of the face and to most deformations, PVHK allows for anonymity and robustness.

REFERENCES

- [1] M. Mahmoudi, S. Sapiro (2009). Three-dimensional point cloud recognition via distributions of geometric distances. In *Graph. Models*.
- [2] M. M. Bronstein and I. Kokkinos (2010). Scale-invariant heat kernel signatures for non-rigid shape recognition. In *CVPR*.
- [3] K. Khoshelham and S. O. Elberink (2012). Accuracy and resolution of kinect depth data for indoor mapping applications. In *Sensors*.
- [4] R. B. Rusu, N. Blodow, and M. Beetz (2009). Fast point feature histograms (fpfh) for 3d registration. In *ICRA*.
- [5] F. Tombari, S. Salti and L. Di Stefan, (2010). Unique signatures of histograms for local surface description. In *ECCV*
- [6] H. Chen and B. Bhanu, (2007). 3D free-form object recognition in range images using local surface patches. In *Pattern Recognition Letters*.
- [7] A. E. Johnson and M. Hebert, (1999). Using Spin Images for Efficient Object Recognition in Cluttered 3D Scenes. In *PAMI*.
- [8] Y. Zhong (2009). Intrinsic shape signatures: A shape descriptor for 3D object recognition. In *ICCV Workshops*.
- [9] R.B. Rusu, G. Bradski, R. Thibaux, and J. Hsu (2010). Fast 3d recognition and pose using the viewpoint feature histogram. In *IROS*.
- [10] R. Osada, T. Funkhouser, B. Chazelle and D. Dobkin, (2002). Shape Distributions. In *ACM Trans. Graph.*
- [11] C. Y. Ip, D. Lapadat, L. Sieger and W. C. Regli, (2002). Using shape distributions to compare solid models. In *SMA '02*.
- [12] W. Wohlkinger and M. Vincze, (2011). Ensemble of shape functions for 3D object classification. In *ROBIO*.
- [13] M. M. Bronstein and I. Kokkinos, (2010). Scale-invariant heat kernel signatures for non-rigid shape recognition. In *CVPR*.
- [14] K. Lai, L. Bo, X. Ren, F. Fox, (2011). Sparse distance learning for object recognition combining RGB and depth information. In *ICRA*.
- [15] M. Blum, J.T. Springenberg, J. Wülfing, and R. Riedmiller, (2012). A learned feature descriptor for object recognition in rgb-d data. In *ICRA*.
- [16] A. Kovnatsky, M. M. Bronstein, A. M. Bronstein, R. Kimmel, (2011) Photometric heat kernel signatures. In *SSVM*.
- [17] J.B. Tenenbaum, V. de Silva and J.C. Langford, (2000). A global geometric framework for nonlinear dimensionality reduction. In *Science*.
- [18] R. B. Rusu, S. Cousins (2011), 3D is here: PCL. In *ICRA*.