HAL
open science

Enhanced Pose Normalization and Matching of Non-Rigid Objects based on Support Vector Machine Modelling

Panagiotis Papadakis

To cite this version:

Panagiotis Papadakis. Enhanced Pose Normalization and Matching of Non-Rigid Objects based on Support Vector Machine Modelling. Pattern Recognition, 2013, 10.1016/j.patcog.2013.06.024 . hal-00844888

HAL Id: hal-00844888

https://inria.hal.science/hal-00844888v1

Submitted on 16 Jul 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Enhanced pose normalization and matching of non-rigid objects based on support vector machine modelling

Panagiotis Papadakis

Auto Agent Laboratory for Cognitive Robotics, Department of Informatics and Systems, University of Rome 'La Sapienza', Italy

A B S T R A C T

The estimation of 3D surface correspondence constitutes a fundamental problem in shape matching and analysis applications. In the presence of non-rigid shape deformations, the ambiguity of surface correspondence increases together with the complexity of registration algorithms.

In this paper, we alleviate this problem by means of 3D pose normalization using One-Class Support Vector Machines (OCSVM). In detail, we show how OCSVM are employed in order to increase the consistency of translation and scale normalization under articulations, extrusions or the presence of outliers. To estimate the relative translation and scale of an object, we use the 3D distribution of points that is modelled by employing OCSVM to estimate the decision surface corresponding to the surface points of the object under a preset tolerance to outliers. By discarding the outliers in the estimation of the object's center and size we compute the canonical pose of the core distribution that is less sensitive to intra-class shape variations. The effectiveness of the proposed method is demonstrated through the increased stability of translation and scale normalization and further justified by improving the precision of content-based 3D object retrieval.

1. Introduction

The increasing availability of 3D content together with the advent of affordable 3D acquisition technology has stimulated research in methods that facilitate the subsequent processing of 3D objects. Within the plurality of applications where 3D objects are used, one of the most commonly encountered problems is pose normalization which decomposes into the normalization of the translation, scale and rotation. Pose normalization is most often performed in applications such as content-based retrieval, thumbnail generation, visualization and modelling where 3D objects need to be positioned in a canonical frame. Pairwise surface registration provides the optimal solution for the shape correspondence problem that is drawing an increasing research interest and has recently been introduced in SHREC (SHape REtrieval Contest) [1,2]. Pose normalization is an alternative approach that provides a single, global solution for each 3D object instance which results in significant efficiency gains at the cost of sub-optimal surface correspondence.

In this paper, we focus on the translation and scale normalization components of the 3D pose normalization problem that have been relatively undervalued in comparison to rotation normalization. However, all three components are equally important to obtain consistent shape correspondence. We demonstrate that the state-of-the-art approaches for translation and scale normalization become insufficient as class-variation increases due to non-rigid shape deformations such as articulation, extrusion or the presence of isolated-outlying parts. To alleviate this problem, we propose the usage of a *novelty detection* algorithm, namely, the One-Class Support Vector Machines (OCSVM) [3,4] as a more robust approach. Using OCSVM we estimate the support of the high-dimensional distribution corresponding to a set of points sampled from the surface of a 3D object and derive the underlying distribution corresponding to the object as the 3D volume that is constrained within the boundaries of the OCSVM decision surface in 3D space. This distribution is then used to derive the relative translation and scale of the object in order to perform a more consistent normalization. This is achieved by lowering the sensitivity of normalization to relatively trivial parts of the object through outlier identification and employing a volume-based approach that is more descriptive compared to solely considering the surface of the object. A significant advantage of the proposed approach is its applicability to *polygon soup* 3D objects in contrast to state-of-the-art methods whose application is constrained to watertight 3D objects [2]. To evaluate the performance, we use a dataset of 3D objects where we introduce various extrusions, articulations or outliers and demonstrate the increased

consistency of intra-class surface correspondence. We complement our experiments by demonstrating the performance gain that is achieved in content-based 3D object retrieval after employing the improved pose normalization process.

This paper complements and extends previous work presented in [5] mainly in the following aspects: (i) a thorough elaboration is provided in support of the formulation of the optimization problem, choice of the kernel, the setting of the parameters and implementation details, (ii) additional qualitative examples are provided from training to evaluation and (iii) the proposed approach is effectively employed to improve the performance of non-rigid 3D shape retrieval in several standard benchmarks.

The remainder of the paper is organized as follows: In Section 2, we discuss related work in the area of translation and scale normalization and outline the limitations of the state-of-the-art approaches in order to motivate the proposed methodology. In Section 3, we formally describe the context in which we employ OCSVM to estimate the distribution of a 3D object that is used to derive the corresponding normalizing translation and scale. In Section 4, we provide a quantitative evaluation that demonstrates the superiority of the proposed method in increasing the consistency of surface correspondence and in turn the performance of content-based retrieval, and finally, in Section 5, we summarize the contributions of this paper.

2. Related work

3D pose normalization accounts for the normalization of the translation, scale and rotation of 3D objects. This step is required in applications where 3D objects need to be positioned in a canonical coordinate frame as, in general, these three characteristics are arbitrarily set. For example, it is necessary in content-based retrieval when shape matching is based on the establishment of correspondences between the surfaces of objects and achieving invariance to isometric or non-isometric transformations.

In the domain of shape matching, the problem of achieving invariance to shape transformations may be viewed from three alternative perspectives. One family of methods concerns the extraction of transformation invariant shape features. As an indicative reference to related work, we may point to the '*Shape-DNA*' signature [6] as well as variants of diffusion and heat kernel descriptors such as those described in [7–10] that have become very popular due to their inherent invariance properties and discriminative power. Another family of methods address the problem of shape matching through pair-wise surface correspondence mainly on the basis of the well known ICP algorithm, that is often adjusted to provide more robust solutions according to the context, e.g. as in [11,12] or [13].

The last family of methods regards the normalization of the 3D pose of objects, in terms of similarity transformations. Among the three normalization problems, it is generally admitted that rotation normalization is the most challenging problem with a significant amount of related work [14–18], where the spatial, orientation or symmetry distribution of the surface has been used to determine the rotation of the frame of the object. However, since rotation normalization is performed on the basis of an appropriately chosen center of an object, translation normalization determines the effectiveness of pose normalization to a greater extent. Therefore, we focus on the translation normalization of 3D objects as well as scale normalization that has been relatively undervalued although it is a difficult problem in a number of cases.

Translation normalization. This problem is most frequently addressed by computing the center of mass of a 3D object and by translating it to make its center coincide with the coordinates origin. For 3D geometric objects consisting of polygons, the center of the object is computed as a weighted average of the total set of the polygon vertices [19], or centroids of the polygons [14], i.e. as the centroid of its surface. For the majority of 3D objects, this method is effective and objects are well normalized with respect to translation. However, there are cases where this method is not appropriate such as articulated 3D objects or objects with isolated-outlying or extruding parts. Such cases are demonstrated in Fig. 1 (a) and (b), respectively.

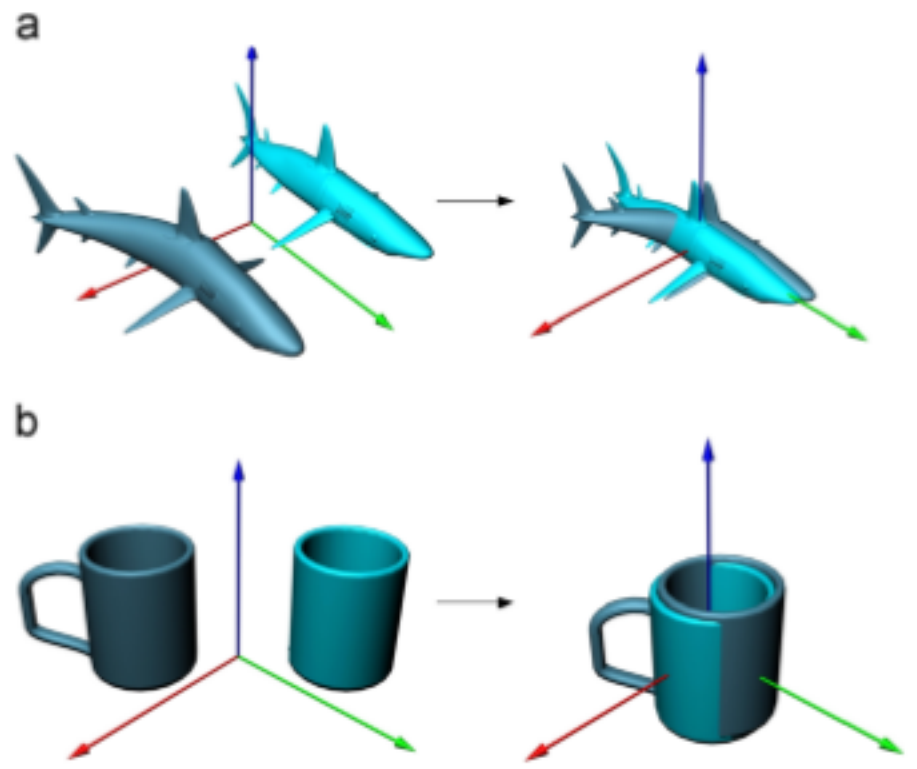

Fig. 1. Translation normalization using the centroid of the surface of the object; (a) example in an object with and without a non-rigid deformation and (b) in an object with and without extruding parts.

Intuitively, we would agree that the centers of the two objects in either case should coincide (in a semantic context). However, by setting the centroid of the surface of an object as its center, we obtain a result that does not agree with our perception. As it is evident from the examples, a small change in the shape of 3D objects can have significant impact on the resulting translation normalization which may in turn negatively affect scale and rotation normalization. Clearly, we expect that the perturbation of the centroid of the surface of an object becomes more pronounced as the amount of articulation or extrusion increases.

In the work of Podolak et al. [16], the authors propose the center of symmetry computed by the Planar Reflective Symmetry Transform as an alternative approach to compute the center of an object. They show that this method is more robust in consistently determining the center of 3D scans of objects, however the performance of the method depends on the symmetry properties of the objects that become less descriptive in the presence of non-rigid transformations.

The smallest enclosing ball of an object [20] has also been considered as a method to determine its center as well as its scale [21]. However, this approach is directly dependent to the outlying parts of an object which makes it the most unstable method as will be demonstrated at Section 4.

Liu and Ramani [18] identified the negative effect of outliers in the rotation normalization component of pose normalization and employed a Random Sample Consensus (RANSAC) [22] approach to find the best model of the object's principal directions. However they did not investigate the effect of outliers in the estimation of the relative translation or scale, that together have a greater impact on the consistency of pose normalization.

Scale normalization. This problem is traditionally addressed by setting the scale of a 3D object to fit within a bounding volume of fixed size (unit cube or unit sphere), or to a size proportional to the

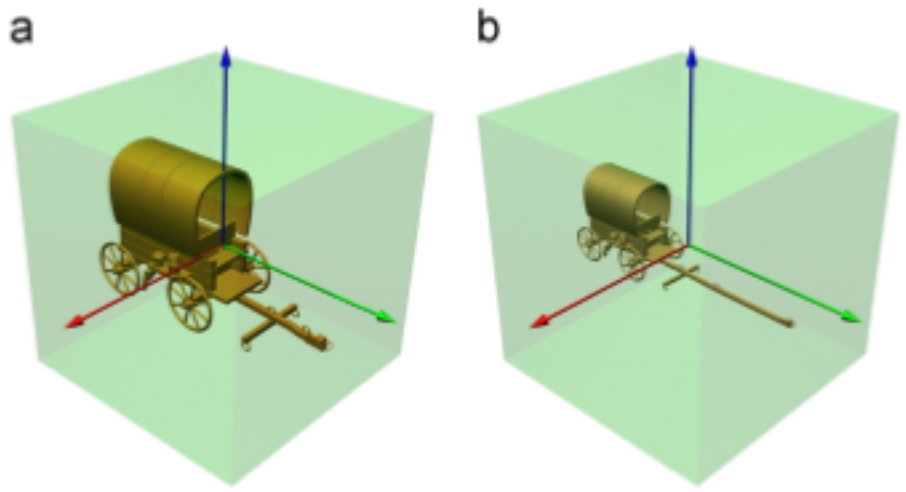

Fig. 2. Scale normalization to the unit cube. The scale normalized version of the 3D object shown in (a) is not consistent in terms of shape correspondence with the scale normalized version of its deformed version with elongated carriage axis in (b).

average distance of the surface of an object from its centroid [15,14].

The performance of the first method is problematic in the presence of outlying parts that will cause an undesired shrinking of the entire object in order to make it fit within the bounding volume. A characteristic example is shown in Fig. 2 where the objects are scaled so as to fit within the unit cube. The object in Fig. 2 (b) is the same in every aspect as the object in Fig. 2 (a) not considering the elongated carriage axis. However, the scale normalized versions of the two objects are significantly different.

Although scaling to a size proportional to the average distance of the surface of an object from its centroid alleviates this problem, it is more a heuristic than a sound methodology. In fact, the consistency of the normalization can only be evaluated by visually examining each 3D object individually. Moreover, the computation of the average distance is still dependent on the entire surface which implies that the effect of trivial outlying parts is not trivial if they are distant from the main cluster of points.

3. Pose normalization using one-class support vector machines

One-class support vector machines (OCSVM) is a *novelty detection* algorithm [3] that estimates the support of a high-dimensional distribution characterised by a set of points labeled as belonging to the distribution. It has been employed for applications that range from classification [23–26] to medical imaging [27], outlier detection [28,29] and surface reconstruction [30], but this is the first time that it is employed to address the translation and scale normalization of 3D objects.

3.1. Distribution estimation of a 3D object

We use OCSVM to estimate a 3D distribution corresponding to the surface of a 3D object that enables a more consistent computation of its relative translation and scale. In detail, we consider the surface of a 3D object as a collection of 3D points that are sampled from the corresponding 3D distribution and we use OCSVM [31] to compute the decision surface that separates this distribution from the residual Euclidean space. The distribution that corresponds to the object is the volume that is constrained within the boundaries of the decision surface.

Let $X=\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_m\}$ denote the set of points that are densely sampled from the surface of a 3D object, where m is the number of points, $\mathbf{x}_i \in T$ and $T \subseteq \mathbb{R}^3$, as shown in Fig. 3. In our experiments we have set the number of sampled points to a fixed value, namely 4000 points, in order to render the proposed approach invariant to varying resolutions and at the same time provide a sufficient number of training points.

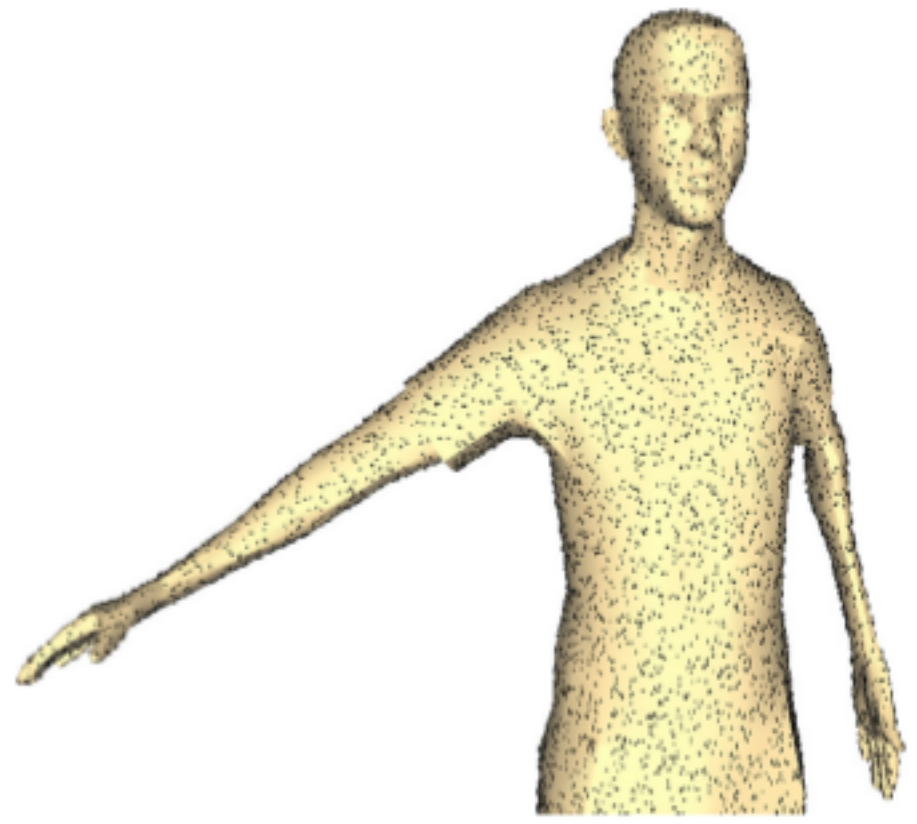

Fig. 3. Points $X=\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_m\}$ sampled from the surface of a 3D object and used for training the OCSVM.

Let Φ be a feature mapping from space T to a dot product space H and $k(\mathbf{x},\mathbf{y})$ a kernel function such that $\langle \Phi(\mathbf{x}),\Phi(\mathbf{x}')\rangle = k(\mathbf{x},\mathbf{x}')$, where $\langle .,. \rangle$ denotes the dot product in H. We want to compute a hyper-sphere with center $\mathbf{c}$ and radius R that contains the maximum number of points $\mathbf{x}_i$ while having the minimum possible radius. This enables identifying parts that are comparatively small to the overall shape and distant from the main cluster of points that could cause a drifting in the computation of the relative translation and scale. Here, the notion of *small* parts of a 3D shape compared to the overall shape can be considered as a portion less than 50% of the overall shape. While there cannot be a general definition of smallness the aforementioned definition could comprise a minimal, reasonable expression.

The computation of the hyper-sphere is formulated as an optimization problem described as follows:

$$\min_{\mathbf{c}\in H, R\in\mathbb{R}} \quad R^2+\frac{1}{\nu m}\sum_{i=1}^{m}\xi_i, \quad \text{subject to} \quad \|\Phi(\mathbf{x}_i)-\mathbf{c}\|^2 \le R^2+\xi_i \tag{1}$$

Here ξ_i denote the "slack" variables that permit the exclusion of outliers from the interior of the hypersphere, i.e. parts that are relatively trivial if they reside far from the primary distribution of points. Slack variables enable dealing with non-rigid shape deformations such as extrusions and articulations or isolated parts of the object that ideally should not affect the determination of the relative translation and scale. The parameter $\nu \in (0,1]$ is used to control the amount of slack, in other words, the tolerance to outliers. By decreasing the value of ν, we increase the tolerance of the algorithm to outliers, while increasing the value of ν we achieve the opposite effect.

To solve the problem, we convert the objective function to its dual form and derive the Lagrangian, from which we get the dual problem:

$$\min_{\alpha} \quad \sum_{i,j=1}^{m}\alpha_i\alpha_j k(\mathbf{x}_i,\mathbf{x}_j) - \sum_{i=1}^{m}\alpha_i k(\mathbf{x}_i,\mathbf{x}_i),$$
$$\text{subject to } 0\le\alpha_i\le\frac{1}{\nu m} \text{ and } \sum_{i=1}^{m}\alpha_i=1 \tag{2}$$

We solve Eq. (2) by employing Sequential Minimal Optimization

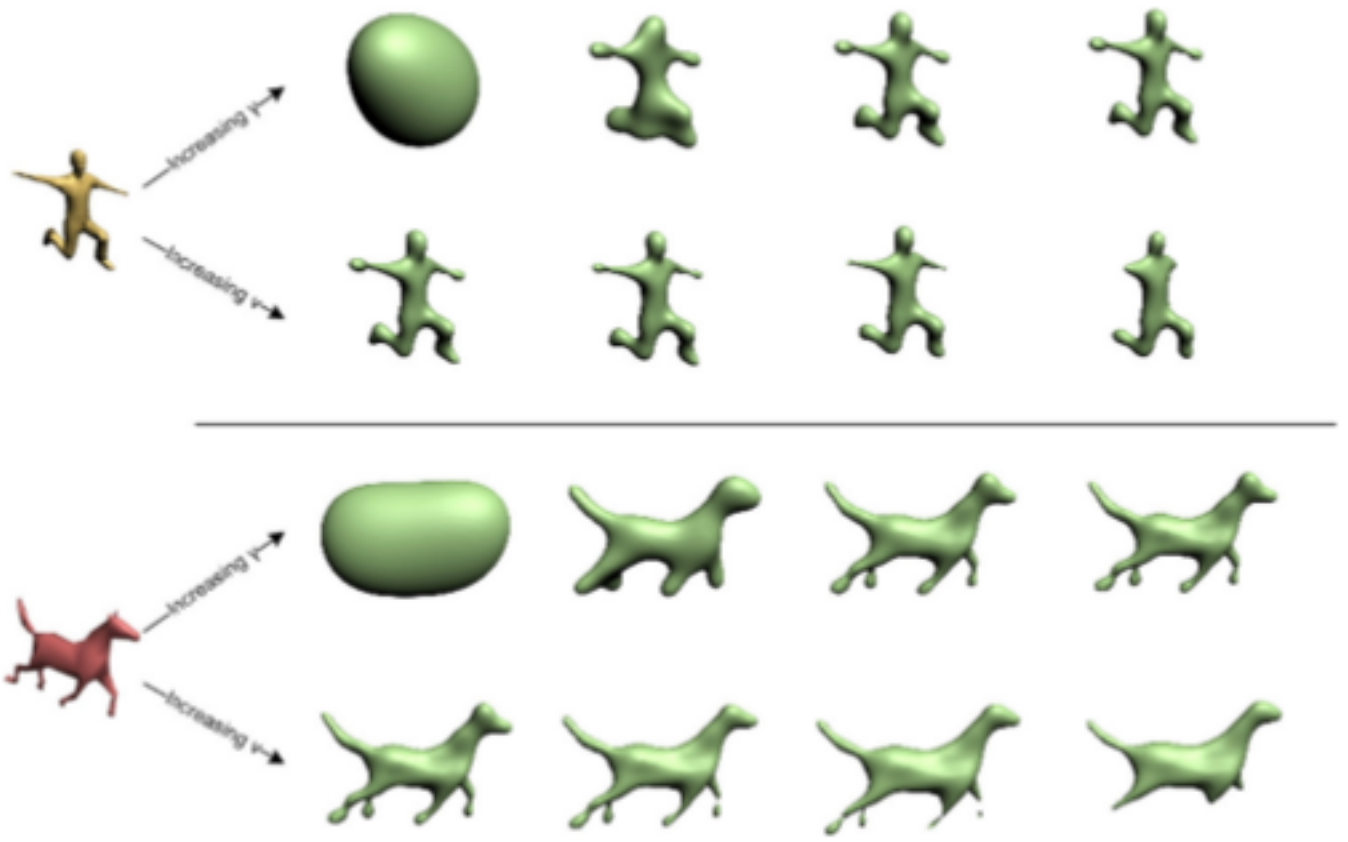

Fig. 4. Example 3D objects and their estimated distribution (depicted in green) using OCSVM for varying γ and ν parameters ($\gamma \in \{1, 15, 31, 43\}$ and $\nu \in \{0.01, 0.2, 0.4, 0.6\}$). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

[32], from which we obtain the coefficients α_i and derive the center $\mathbf{c}$ of the hyper-sphere. The decision function that models the distribution of points is determined by $f_s(\mathbf{x}) = \mathrm{sgn}(f(\mathbf{x}))$ where sgn is the sign operator, giving $+1$ if the input is positive, -1 if it is negative and 0 otherwise, and $f(\mathbf{x})$ is given by

$$\begin{aligned} f(\mathbf{x}) &= R^2 - (\mathbf{c} - \Phi(\mathbf{x}))^2 = R^2 - \left[\sum_{i=1}^{m} \alpha_i \Phi(\mathbf{x}_i)\right]^2 \\ &\quad + 2\sum_{i=1}^{m} \alpha_i \Phi(\mathbf{x}_i)\Phi(\mathbf{x}) - [\Phi(\mathbf{x})]^2 \\ &= R^2 - \sum_{i,j=1}^{m} \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) + 2\sum_{i=1}^{m} \alpha_i k(\mathbf{x}_i, \mathbf{x}) - k(\mathbf{x}, \mathbf{x}) \end{aligned} \tag{3}$$

If $f_s(\mathbf{x}) = +1$ then $\Phi(\mathbf{x})$ lies inside the hypershere and $\mathbf{x}$ is considered to belong to the estimated distribution, while if $f_s(\mathbf{x}) = -1$, $\Phi(\mathbf{x})$ lies outside the hypersphere and $\mathbf{x}$ is considered not to belong to it. If $f_s(\mathbf{x}) = 0$ then $\mathbf{x}$ lies on the decision surface, as in the case of support vectors. Hence, the 3D object is now modelled as the sublevel set of f_s wherein $f_s(\mathbf{x}) \le 0$.

In general, OCSVM can be employed using different kernels and settings of parameters that determine the final shape of the decision surface in the observation space. Following the results presented in [3] and [33], we choose the Gaussian kernel $k_G(\mathbf{a}, \mathbf{b}) = e^{-\gamma\|\mathbf{a}-\mathbf{b}\|^2}$, where $\mathbf{a}, \mathbf{b} \in T$, since it bears a number of desirable properties:

- By the definition of the Gaussian kernel, the product of all mapped points is always positive which implies that they all lie on the same orthant of the projection space. The latter, combined with the fact that all mapped points have unit length, ensures that they are separable from the origin in the projection space [34] and hence there is always a solution.
- In comparison to the polynomial kernel, the Gaussian kernel produces decision surfaces that are generally more tight to the data points [33].
- The problem of estimating the minimum bounding hyper-sphere becomes mathematically equivalent to the problem of estimating the hyperplane that separates the data from the origin with maximum margin, essentially giving the same shape of decision surface in the original space [34].
- Through proper moderation of the σ parameter of the kernel and the slack parameter ν, we have direct guaranteed control on the number of support vectors lying on the decision surface, hence, controlling the complexity of the iso-surface.

Using the Gaussian kernel the OCSVM algorithm is parameterized by the γ variable of the kernel and the variable ν that controls the slack. The γ variable controls the shape of the decision surface in 3D space in terms of its similarity with respect to the sphere and the underlying surface of the object. In particular, as $\gamma \to 0$ the decision surface approaches the shape of the sphere, while, as γ increases, the decision surface fits better to the surface of the object. This can be better understood by considering that γ is inversely proportional to the σ^2 of the Gaussian kernel, thus, when $\sigma \to 0$ the uncertainty is decreased and the decision surface tends to fit the surface of the object. The ν variable controls the amount of points from the surface of the object that will reside within the decision surface and it can be used to improve the robustness of the algorithm with respect to articulated, extruding or outlying parts. As $\nu \to 0$ outliers are minimized while the opposite is true as $\nu \to 1$. In Fig. 4 we demonstrate the effect of different γ and ν settings in estimating the distribution of 3D objects.

3.2. Translation and scale normalization

OCSVM methodology applied to translation normalization lets us overcome the limitations of conventional approaches that cannot face increasing intra-class variation of 3D objects. The insufficiency is due to the assumption that the center of a 3D object is the center of mass of the surface of the object. This approach is not robust when non-rigid shape deformations are applied to 3D objects, such as articulations, extrusions or the presence of outlying parts (see Fig. 1) where the center of mass of the surface can be considerably perturbed.

The solution that we propose provides an alternative method for computing the center of an object that is less sensitive to the aforementioned scenarios. More specifically, an object center is taken to be the centroid $\mathbf{c}_m$ of the distribution estimated according to the OCSVM method, and whose volume V is within the boundaries of the decision function $f_s(\mathbf{x})$ in the observation space.

The centroid of the distribution is computed as

$$\mathbf{c}_m = \frac{1}{V} \iiint_{f_s(\mathbf{p})\geq 0} \mathbf{p}\, dx\, dy\, dz \tag{4}$$

where $\mathbf{p} = (x, y, z) \in T$ and $V = \iiint_{f_s(\mathbf{p})\geq 0} dx\, dy\, dz$.

The motivation behind the proposed approach for computing the object center is two-fold. First, by using OCSVM we are able to decrease the sensitivity to extruding-outlying parts by appropriately setting the slack, i.e. the ν variable. And second, we adopt a volume-based approach for computing the centroid which is more stable than only considering the surface area of the object. This is achieved by considering not only the decision surface, i.e. where $f_s(\mathbf{x}) = 0$ but also the space where $f_s(\mathbf{x}) = +1$. Furthermore, this enables the computation of a volumetric center even for objects that are not watertight.

After translation normalization, we can use the computed distribution that corresponds to the 3D object to perform scale normalization. To normalize the scale we estimate the average distance d_m of the distribution of the object from its center and scale the object so that this distance is unit. The average distance d_m is computed as

$$d_m = \frac{1}{V} \iiint_{f_s(\mathbf{p})\geq 0} \|\mathbf{p} - \mathbf{c}_m\| dx\, dy\, dz \tag{5}$$

where $\|.\|$ denotes the L_2 norm.

The resulting scale normalization will be more consistent as by employing OCSVM with a preset tolerance to outliers we succeed in lowering the sensitivity of the normalization to trivial parts of the objects. Furthermore, we directly benefit from the more consistent determination of the center of an object using OCSVM since the average distance is computed with respect to this point.

In practice, Eqs. (4)–(5) are evaluated over a finite uniformly distributed set of points $\mathbf{p}$ that form a 3D $s \times s \times s$ voxel grid that incorporates the 3D object, where s is the underlying resolution. The function $f_s(\mathbf{p})$ is evaluated over the whole 3D grid to identify the voxels that belong to OCSVM 3D distribution which are then used to compute the center $\mathbf{c}_m$ and the average distance d_m. Determining a particular value for s depends on the desired level of detail that is needed to model a 3D object that should be balanced to the required computation time, whose complexity is $O(s^2)$ (in our experiments setting $s = 64$ appeared as a satisfactory trade-off).

4. Results

In this section we evaluate the performance of the proposed methodology in terms of the consistency of pose normalization in the presence of outliers or when non-rigid transformations such as articulations and extrusions are applied to a variety of 3D objects. Furthermore, we demonstrate the performance gain that is attained in content-based 3D object retrieval after employing the improved pose normalization process.

4.1. Dataset for training and evaluation of translation and scale normalization

To evaluate the performance of pose normalization, we have generated a set of categorized 3D objects (Fig. 5) that contains 120 generic 3D objects in total, evenly distributed in 12 classes, namely, *human*, *dolphin*, *chair*, *tree*, *pot*, *desktop computer*, *hand*, *dog*, *guitar*, *piano*, *spider* and *horse*.

Each class of the set was created using a reference 3D object. To build the dataset, we introduce various non-rigid deformations to the reference object of each class and generate various articulations, extrusions or add-remove outlying parts. By applying such modifications to the reference object, we create new instances of the same object which altogether constitute a particular class. Since the geometry of the generated objects is different than the geometry of the respective reference object, we expect that the geometric center of the objects should also be altered. However, we do not apply any *similarity transformation* to the generated 3D objects, therefore, the objects of the same class are pose normalized with respect to their core part. This is required in order to evaluate pose normalization and to our knowledge there is no publicly available dataset that satisfies this requirement, therefore standard databases such as [35,36] or [37] cannot be used for this evaluation.

We should note here, that the effect of the γ parameter, unlike the ν parameter, is related to the scale of the objects. For example, employing OCSVM with $\gamma = C$ to a particular 3D object would give the same shape of decision surface as employing OCSVM with $\gamma = 4C$ to the same object at half its size. Therefore, prior to the generation of the dataset, we have normalized the scale of the reference 3D objects so that they fit within the unit cube in order to ensure a consistent evaluation of the performance and that no particular class will bias the results. The initial scaling could have also been avoided without any implication on the results, in which case, the values of the optimal value for the γ parameter of the Gaussian kernel would have been obtained proportionally to a constant rather than obtaining an absolute value. Despite the fact that the final results would have been equivalent, however, we choose not to give relative values for the γ parameter but absolute, since this allows our implementation to be replicated.

4.2. Evaluation of translation and scale normalization

By comparing the standard deviation for the centroid of the surface of an object and the centroid of the distribution of the object as computed by OCSVM, in each class of the dataset, we can evaluate the performance of OCSVM under different settings of the γ parameter of the Gaussian kernel and the slack ν. With respect to the γ parameter, we have evaluated the performance for $\gamma \in (0, 50]$. The value of the upper limit was determined considering the effect of overfitting the estimated distribution to the points of the 3D object.

In Fig. 6(a) we demonstrate how the standard deviation of the center of the distribution computed using OCSVM varies in relation to the γ parameter and compare it against the standard deviation of the centroid of the surface of an object. The performance is averaged over all classes. This evaluation demonstrates the increased consistency of translation normalization achieved by OCSVM solely due to the determination of the center by a volume-based approach rather than a surface-based approach, thus the slack is set to $\nu = 0$ and the complete 3D object surface is taken into account. The diagram shows that the perturbation of the center of the OCSVM distribution initially decreases rapidly as γ increases and finally converges to a global minimum. The point where the curve corresponding to OCSVM is minimized in Fig. 6 (a), i.e. the point where the average perturbation of the center is minimum, is for $\gamma = 47$.

From the definition of the Gaussian kernel, it can be easily seen that regular increments of the γ parameter do not have a linear effect on the estimation of the kernel value and hence on the estimation of the decision surface. In other words, as γ increases, it is more and more difficult to have a noticeable change in the results (as can also be observed in Fig. 4). In view of this result, a global minimum is reached in the sense that for $\gamma > 47$ the change in the standard deviation of the centroid is smaller than a very small threshold ϵ and fluctuations in the performance could appear for significantly higher values of γ.

We interpret this point as the best trade-off between precisely characterizing the surface of an object through the OCSVM distribution

Fig. 5. The dataset of 3D objects used for evaluating the performance of translation and scale normalization. Each column corresponds to a class of 3D objects created by generating articulations, extrusions or adding-removing outlying parts.

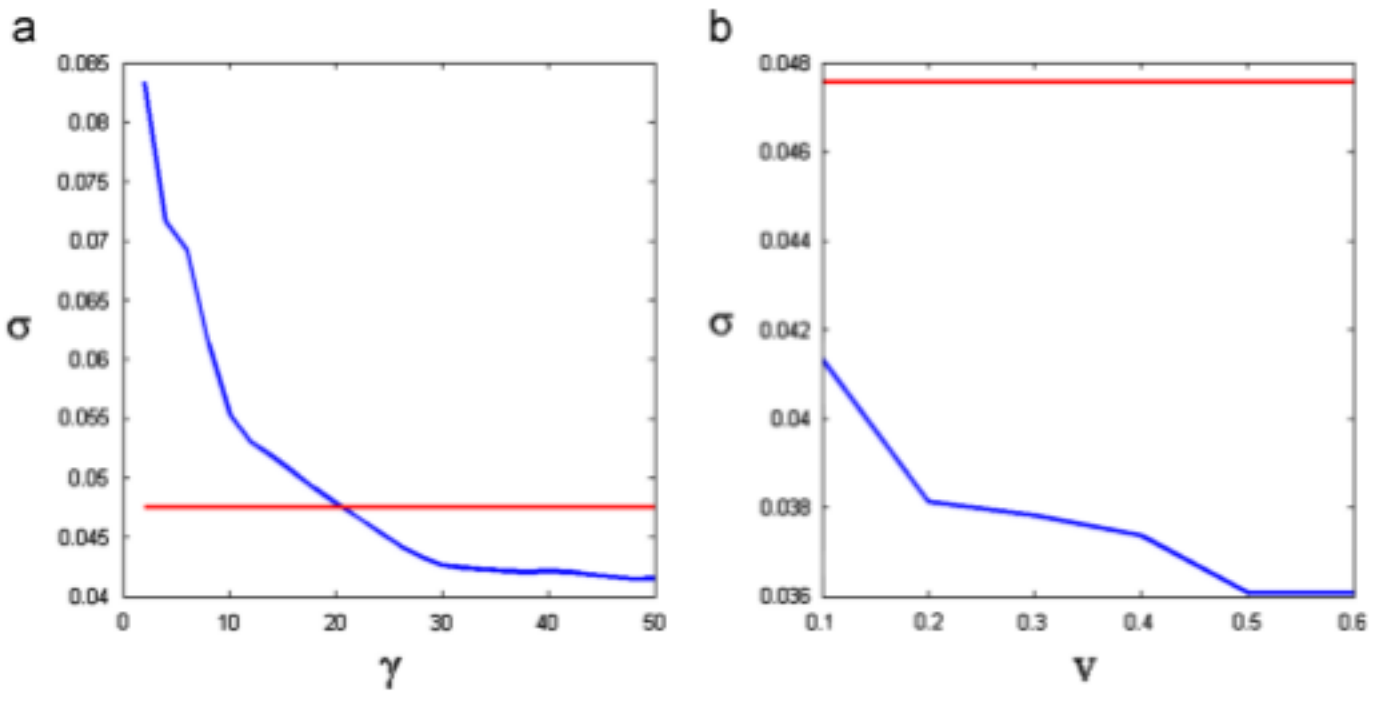

Fig. 6. Average standard deviation of the centroid of an object computed using OCSVM (blue curve) for (a) $\gamma \in (0, 50]$ and $\nu = 0$ and (b) $\nu \in [0.1, 0.6]$ and $\gamma = 47$, compared to the standard deviation of the centroid of the complete surface of an object (red curve). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

and overfitting the distribution to the corresponding surface. Indirectly, Fig. 6(a) also provides an evaluation for the consistency of translation normalization by translating an object to the center of its minimum enclosing sphere (i.e. when $\gamma \to 0$). It is evident that this method exhibits the worst performance compared to the surface centroid-based approach and the proposed approach using OCSVM.

After fixing $\gamma = 47$ we evaluate the performance of OCSVM translation normalization under different settings of the slack and

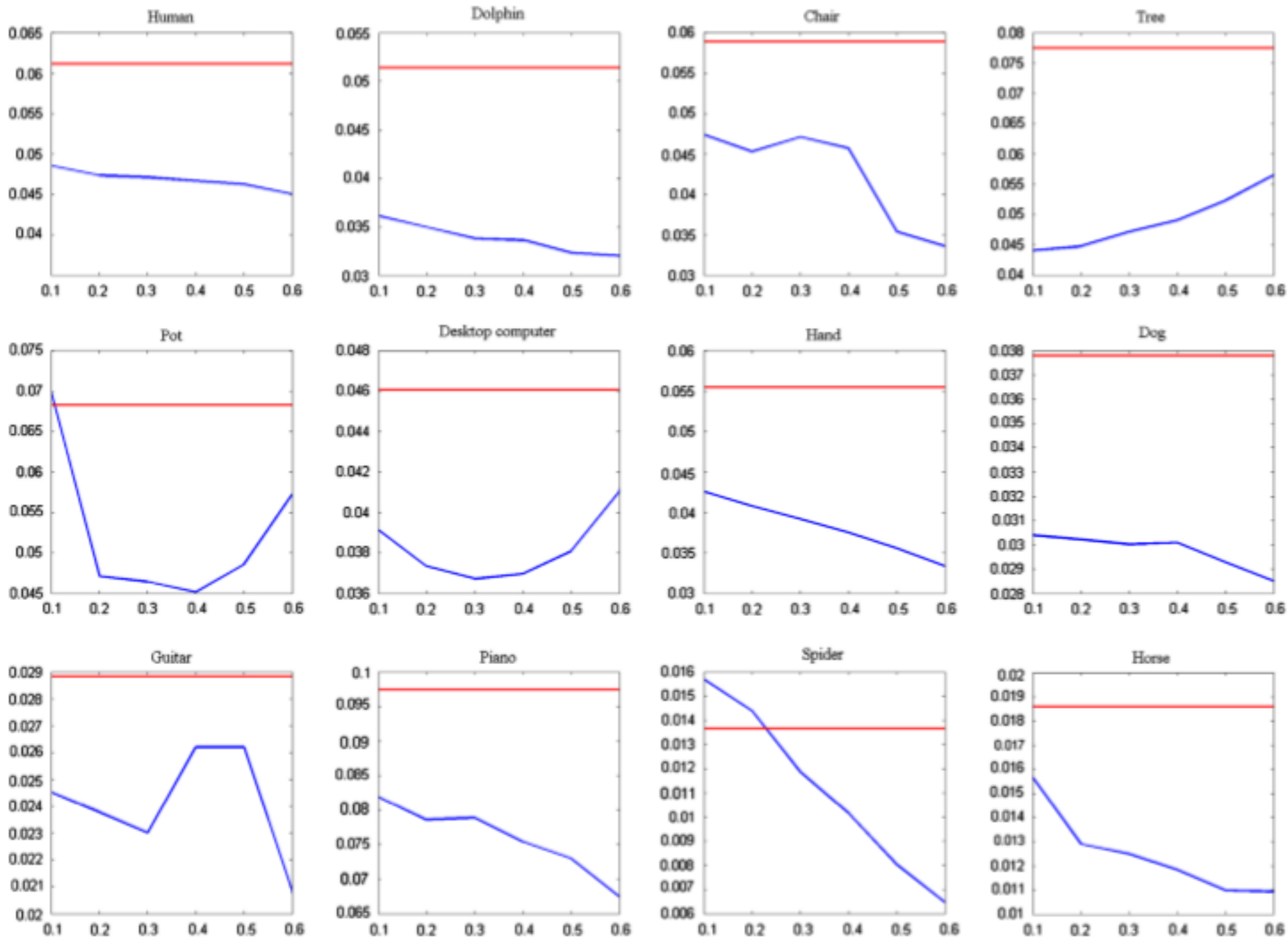

Fig. 7. Average standard deviation of the centroid of an object computed using OCSVM (blue curve) for $\nu \in [0.1, 0.6]$ and fixed $\gamma = 47$, compared to the standard deviation of the centroid of the complete surface of a 3D object (red line). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

in particular for $\nu \in [0.1, 0.6]$, as shown in Fig. 6 (b). The value of the upper limit implies that this is the boundary that we are using to detect outlying parts that should not be considered in the computation of the center.

This evaluation demonstrates the increased consistency of translation normalization that is achieved by OCSVM due to the determination of the center by discarding outlying parts rather than considering the complete 3D surface. The point where the curve corresponding to OCSVM is minimized in Fig. 6 (b), i.e. the point where the average perturbation of the estimated center is minimum, is for $\nu = 0.5$, while for larger values it gradually increases. We interpret this point as the best trade-off between discarding outliers that should not be considered in the center computation and at the same time not discarding the important parts. Apart from the experimental evidence for the optimal parameter setting, this result is also theoretically consistent. In detail, according to the ν property of SVMs, namely, the slack-property, the value of ν is an upper bound on the percentage of outliers in the final distribution estimation after solving the optimization problem. In other words, the experimentally derived optimal setting of $\nu = 0.5$ essentially declares that there will be at most 50% of the overall shape that will be considered as outlier and discarded from the core distribution. Therefore, what we experimentally observe as a deterioration of performance as ν approaches 1 has a theoretical explanation which is that the important, core part of the 3D shape starts to be discarded too.

In Fig. 7, we evaluate the performance within each class of the dataset. Within classes where objects have a well defined core part (human, dolphin, chair, vase, computer, hand, dog, guitar, piano, spider and horse) as well as relatively smaller articulated-outlying parts, we observe that it is meaningful to introduce the slack in the OCSVM distribution estimation and discard a certain proportion of the surface of an object as being an outlier in order to determine its center. An exception is the tree class where the determination of the core part is more ambiguous. Here we observe the inverse behaviour since OCSVM may erroneously discard important parts of an object that are not outliers. In this case, the perturbation of the center increases as meaningful parts are not taken into account, however, the performance gain that is attained due to the volume-based approach for computing the center mostly compensates for this insufficiency. Generally, the performance of the proposed approach is stable for 3D objects where a core part can be identified and the remaining parts of the objects can be discarded up to a predefined ratio.

By setting $\gamma = 47$ and $\nu = 0.5$ the average standard deviation of the OCSVM center of an object is equal to $\sigma_{tr,OCSVM} = 0.0361$ while the average standard deviation of the centroid of the surface of an object is $\sigma_{tr,centroid} = 0.0475$. In other words, we have lowered the

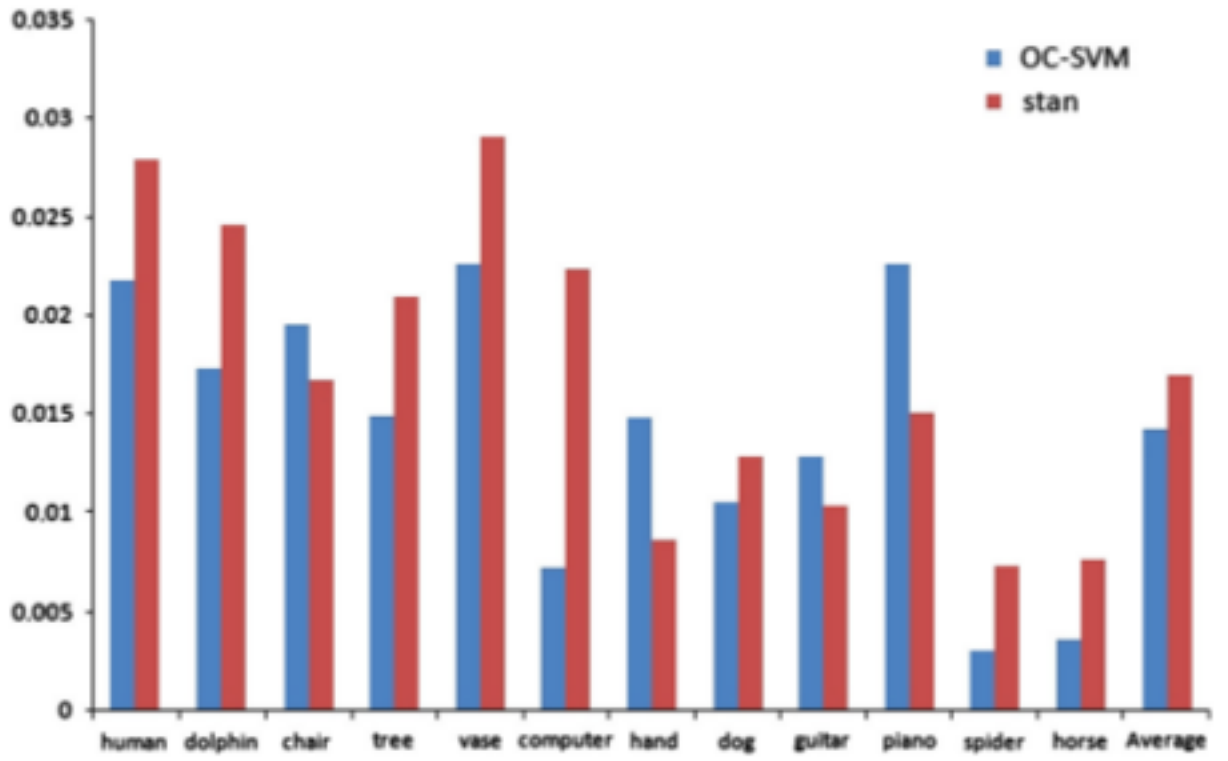

Fig. 8. Evaluation of scale normalization by measuring the standard deviation (vertical axis) of the average distance from the centroid of the distribution using OCSVM and the average distance from the centroid of the surface of the objects (stan).

Fig. 9. Comparison of translation and scale normalization using OCSVM (top row) against using the centroid and average distance of the surface (bottom row).

average perturbation of the center for objects that belong to the same class by 24%. This performance gain does not only improve translation normalization, but also scale normalization which is based on measuring distances from the center of the object. Analogously, we expect that rotation normalization methods that involve computations with respect to the center of an object would benefit as well.

To compare the performance of OCSVM scale normalization against the method that uses the average distance of the surface from its centroid, we measure the standard deviation of the distance for each class and the overall performance is taken by averaging over all classes (Fig. 8).

The results demonstrate that the proposed method tends to be more robust than the standard approach since on average $\sigma_{scale,OCSVM} = 0.01425$ and $\sigma_{scale,stan} = 0.01699$. This means that we have achieved to lower the perturbation of the average distance computation by about 16.1%. The consistency of scale normalization when using OCSVM is higher in 8 out of the 12 classes. The inferior performance in the other 4 classes can be partly attributed to the fact that for scale normalization using OCSVM we set the parameters ν and γ to those values that gave the best results for translation normalization. Although this is not a strict a requirement, we adopt this approach in order to reduce the complexity of the total normalization process to a single run of the OCSVM algorithm. Alternatively, we could determine the values of ν and γ based on the combined performance of translation and scale normalization.

In Fig. 9 we show the result of both translation and scale normalization for pairs of objects of the same class in order to provide a qualitative evaluation of the proposed approach, in terms of shape correspondence.

4.3. Evaluation in content-based 3D object retrieval

We next present the results of content-based 3D object retrieval experiments that demonstrate the performance gain that is attained when the proposed pose normalization method is adopted. For this evaluation, we have employed a contemporary 3D shape descriptor called PANORAMA [38] that attains state-of-the-art discriminative power in real time as far as generic 3D shape matching is concerned. Among other, the advantage of this method is that its application is not constrained to watertight 3D meshes, hence it can be readily applied to *polygon soup*3D objects.

The PANORAMA descriptor uses a set of panoramic views of a 3D object that capture both the spatial position and orientation of an object's surface. The panoramic views are obtained by projecting the object to the lateral surfaces of three cylinders that are centered at the centroid of the object, scaled to a multiple of the object's relative size and aligned to the object's principal axes and for each projection the corresponding 2D Discrete Fourier Transform as well as 2D Discrete Wavelet Transform are employed.

The principal axes are determined by employing a hybrid rotation normalization scheme that uses the Continuous PCA (CPCA) and Normals PCA (NPCA), as introduced in [15]. In short, CPCA is employed in the space of 3D coordinates of the model's surface, while NPCA is employed on the 3D orientation of a model's surface. In the hybrid scheme, both methods are used resulting in two alternative rotation normalized versions of the same 3D model and give two sets of features. When matching is performed within a pair of objects, two comparisons are computed, the first using the CPCA-based and the second using the NPCA-based set of features. The comparison which gives the minimum distance eventually gives the final distance score between the two models.

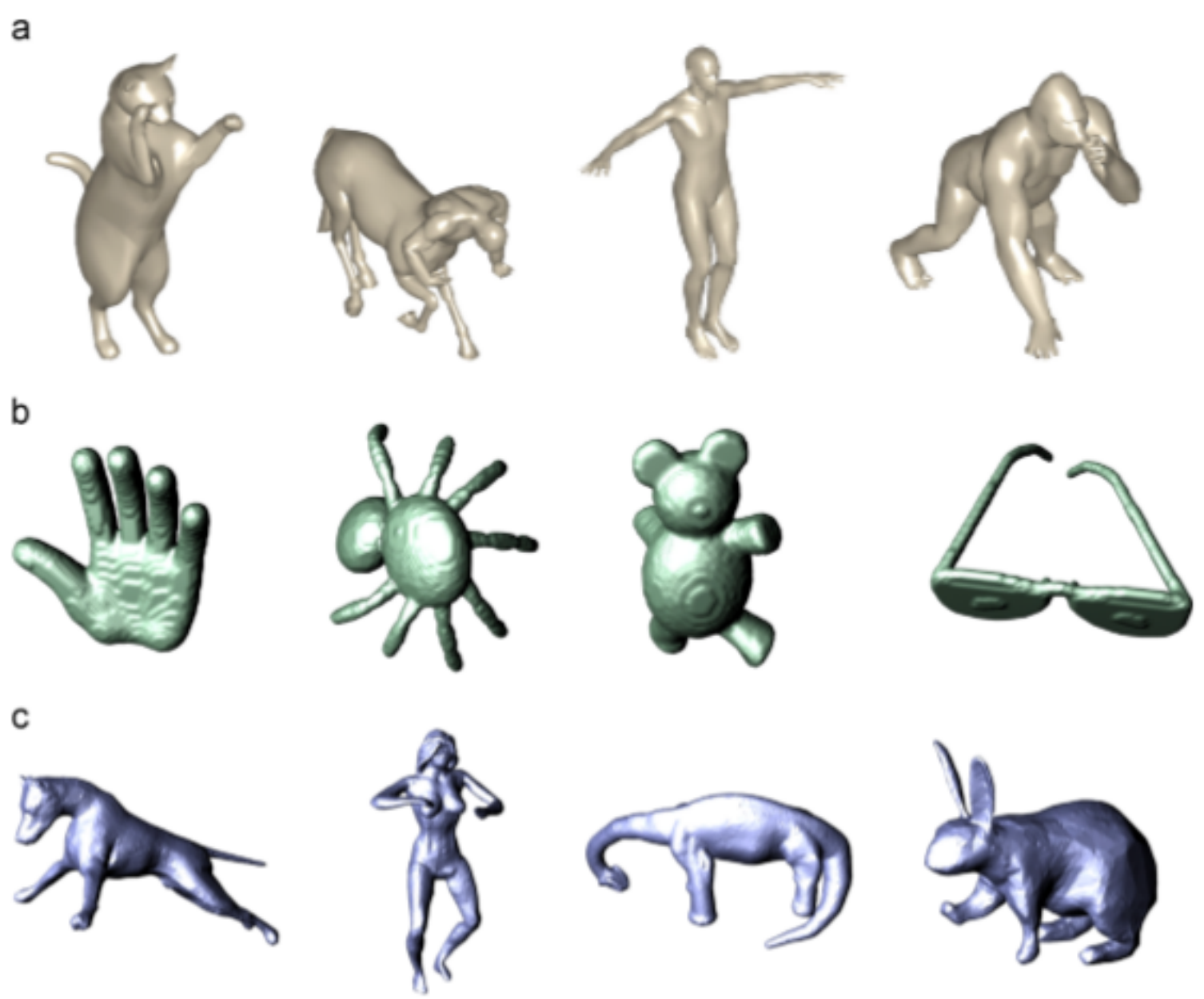

Fig. 10. Instances of 3D objects from (a) the NRW dataset [39], (b) the SHREC-NR10 [40] dataset and (c) the SHREC-NR11 [41] dataset.

For evaluating the performance of content-based 3D object retrieval under non-rigid deformations we have used three standard benchmarks, namely, the *Non-Rigid World* (NRW) dataset [39], the *SHREC Non-Rigid 3D Models dataset 2010* (SHREC-NR10) [40] and the *SHREC 2011 Non Rigid 3D Watertight Meshes* (SHREC-NR11) [41]. The NRW dataset contains in total 148 articulated objects distributed in 12 classes from which we have excluded the *wolves* and *shark* classes that together comprise 4 objects and therefore are of little statistical importance. The SHREC-NR10 dataset contains 200 models distributed over 10 classes while SHREC-NR11 contains 600 models distributed over 30 classes. All these datasets are particularly suited for evaluating the robustness of our pose normalization method under non-rigid shape deformations, since they are characterized by a high degree of non-rigid deformations and hence intra-class variation. In Fig. 10 a set of example 3D objects are shown from each dataset.

To evaluate the performance we use precision-recall diagrams. Recall is the ratio of relevant to the query retrieved models to the total number of relevant models while precision is the ratio of relevant to the query retrieved models to the number of retrieved models. The evaluations were performed by using each model of a dataset as a query in the remaining set of models and computing the average precision-recall performance over all the models.

In Fig. 11, we demonstrate the increase in retrieval precision that is achieved when the proposed translation and scale normalization approach is adopted (PANORAMA-OCSVM) compared to translating to the centroid of the complete 3D surface and using the average distance of the complete surface from its centroid to compute the relative scale (PANORAMA). Furthermore, we have evaluated the performance in terms of the scores of each run with respect to the nearest neighbour (NN), first-tier (FT) and second-tier (ST) evaluation measures that are commonly used in content-based 3D object retrieval experiments (we refer the reader to [35] for the exact formulation of these measures).

Table 1 shows the attained scores in all three datasets and summarizes the overall performance by weighing the contribution of each dataset according to its cardinality.

The results show a noticeable increase in the overall performance when the proposed OCSVM-based pose normalization method precedes the extraction of the PANORAMA descriptor. By taking into account the cardinality of each dataset, we obtain that the overall average increase is in the order of 1.3%, 2% and 2.6% for the NN, FT and ST evaluation measures, respectively. This implies that the proposed method can be employed to increase the discriminative power of any generic 3D shape descriptor that requires the normalization of the object's pose, since the OCSVM-based pose normalization can be readily applied to any kind of 3D object representation. These performance gains are attained based only on the training being performed in the dataset shown in Fig. 5 (containing 120 models), where the sensitivity of OCSVM-based pose normalization was evaluated, against the standard approach. If the setting of the best parameters was tuned according to the performance of content-based retrieval and within a bigger corps of data, higher performance gains can be achieved.

As far as efficiency is concerned, the time required for the OCSVM based pose normalization has been measured for each instance within all three datasets, yielding an average

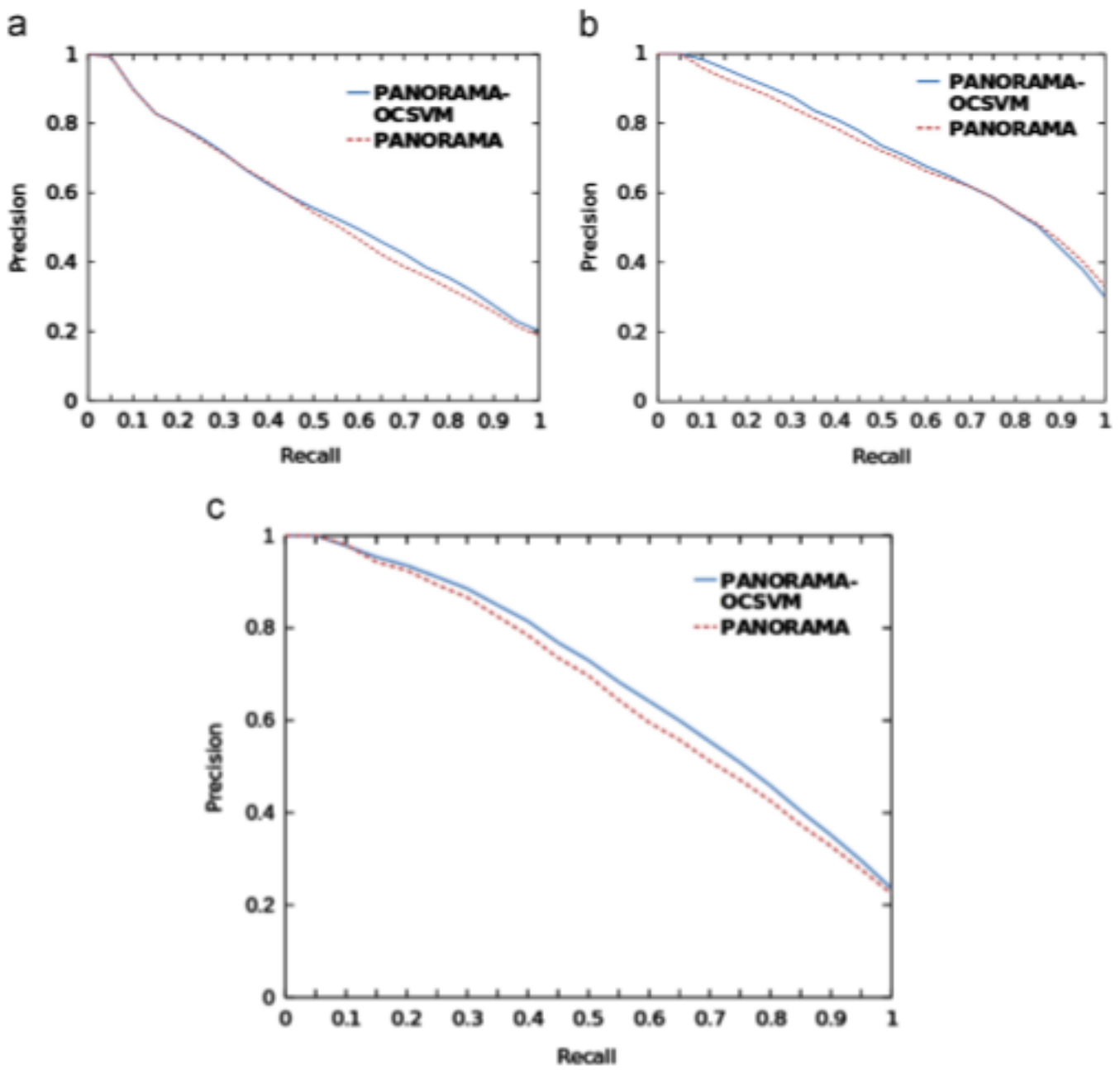

Fig. 11. Precision-recall diagram of the PANORAMA descriptor within (a) the NRW dataset, (b) the SHREC-NR10 and (c) the SHREC-NR11 showing the performance when the proposed pose normalization method is adopted (PANORAMA-OCSVM) against the original descriptor (PANORAMA).

Table 1
Quantitative measures scores of the Panorama-OCSVM and Panorama for each evaluation dataset individually and over all datasets.

Method	NRW			SHREC-NR10			SHREC-NR11		
	NN (%)	**FT** (%)	**ST** (%)	**NN** (%)	**FT** (%)	**ST** (%)	**NN** (%)	**FT** (%)	**ST** (%)
Panorama-OCSVM	75.5	45.0	61.0	97.0	62.0	71.5	95.5	61.5	71.5
Panorama	73.6	43.0	58.7	92.0	61.5	71.5	95.5	58.5	68.0

Method	TOTAL		
	NN (%)	**FT** (%)	**ST** (%)
Panorama-OCSVM	**92.7**	**58.8**	**69.9**
Panorama	91.4	56.8	67.3

computation time of 2.3 s $\pm$0.2 s, using a contemporary computer equipped with an Intel Core i7 CPU 860 at 2.8 Ghz. If needed, the computation time can be considerably decreased with a simple adjustment, namely, by using a precomputed array of kernel values, which is possible according to the formulation of the proposed pose normalization technique. In detail, since we always solve the optimization problem within a constrained space (unit cube) for all objects, one may precompute the Gaussian kernel values for a fixed space interval and use the norm of the difference $\|\mathbf{x}-\mathbf{y}\|$ between two vectors $\mathbf{x}$ and $\mathbf{y}$ as an index into this array. This step can speed up both the training as well as the prediction stage.

5. Conclusions

In this paper, we have presented a method that robustly addresses the problem of translation and scale normalization of 3D objects under non-rigid shape transformations such as articulations, extrusions or the presence of outliers. The proposed method achieves a more consistent surface correspondence between 3D objects compared to the prominent pose normalization approach that considers the entire 3D object surface in order to compute the relative translation and scale which renders it sensitive to the aforementioned shape deformations.

The proposed method has been quantitatively and qualitatively evaluated and found superior in terms of intra-class surface correspondence as well as shape matching precision for content-based 3D object retrieval. Along with the performance gain that is achieved, the proposed pose normalization method is beneficial compared to non-rigid shape matching methods whose application is constrained to manifold 3D meshes or of certain genus, as it is applicable to *polygon soup* 3D meshes and therefore can be readily employed to increase the performance of any shape signature for generic 3D objects.

Conflict of Interest

None declared.

Acknowledgments

We would like to thank the reviewers for their invaluable feedback both in terms of improving the quality of the manuscript as well as for stimulating and inspiring our future work.

References

[1] A.M. Bronstein, M.M. Bronstein, U. Castellani, A. Dubrovina, L.J. Guibas, R.P. Horaud, R. Kimmel, D. Knossow, E. von Lavante, D. Mateus, M. Ovsjanikov, A. Sharma, Shrec 2010: robust correspondence benchmark, in: Proceedings of the EUROGRAPHICS Workshop on 3D Object Retrieval, 2010.

[2] Z. Lian, A. Godil, B. Bustos, M. Daoudi, J. Hermans, S. Kawamura, Y. Kurita, G. Lavoue, H. Nguyen, R. Ohbuchi, Y. Ohkita, Y. Ohishi, F. Porikli, M. Reuter, I. Sipiran, D. Smeets, P. Suetens, H. Tabia, D. Vandermeulen, Shrec'11 track: shape retrieval on non-rigid 3d watertight meshes, in: Eurographics Workshop on 3D Object Retrieval, 2011.

[3] B. Scholkop, J. Platt, J. Shawe-Taylor, A. Smola, R. Williamson, Estimating the support of a high-dimensional distribution, Neural Computation 13 (2001) 1443–1471.

[4] B. Scholkopf, A. Smola, Learning with Kernels, The MIT Press, 2002.

[5] P. Papadakis, F. Pirri, Consistent pose normalization of non-rigid shapes using one-class support vector machines, in: IEEE International Conference of Computer Vision and Pattern Recognition Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment (NORDIA), 2011.

[6] M. Reuter, F. Wolter, N. Peinecke, Laplace-beltrami spectra as 'shape-dna' of surfaces and solids, Computer Aided Design 38 (2006) 342–366.

[7] J. Sun, M. Ovsjanikov, L. Guibas, A concise and provably informative multi-scale signature based on heat diffusion, in: Eurographics Symposium on Geometry Processing, 2009, pp. 1383–1392.

[8] K. Gebal, J. Baerentzen, H. Aanaes, R. Larsen, Shape analysis using the auto diffusion function, in: Eurographics Symposium on Geometry Processing, 2009, pp. 1405–1413.

[9] M.M. Bronstein, I. Kokkinos, Scale-invariant heat kernel signatures for non-rigid shape recognition, in: IEEE International Conference on Computer Vision and Pattern Recognition, 2010, pp. 1704–1711.

[10] D. Raviv, M. Bronstein, A. Bronstein, R. Kimmel, N. Sochen, Affine-invariant diffusion geometry for the analysis of deformable 3D shapes, in: IEEE International Conference on Computer Vision and Pattern Recognition, 2011, pp. 2361–2367.

[11] A. Bronstein, M. Bronstein, Regularized partial matching of rigid shapes, in: Proceedings of the 10th European Conference on Computer Vision, 2008, pp. 143–154.

[12] A. Bronstein, M. Bronstein, Not only size matters: regularized partial matching of nonrigid shapes, in: IEEE International Conference on Computer Vision and Pattern Recognition Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment, 2008, pp. 1–6.

[13] F. Pomerleau, S. Magnenat, F. Colas, M. Liu, R. Siegwart, Tracking a depth camera: parameter exploration for fast ICP, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, 2011.

[14] D. V. Vranic, 3D model retrieval, Ph.D. Thesis, University of Leipzig, 2004.

[15] P. Papadakis, I. Pratikakis, S. Perantonis, T. Theoharis, Efficient 3D shape matching and retrieval using a concrete radialized spherical projection representation, Pattern Recognition 40 (9) (2007) 2437–2452.

[16] J. Podolak, P. Shilane, A.G.S. Rusinkiewicz, T. Funkhouser, A planar-reflective symmetry transform for 3d shapes, ACM Transactions on Graphics 25 (2006) 549–559.

[17] P. Papadakis, Content-based 3D Model Retrieval Considering the Users Relevance Feedback, Ph.D. Thesis, University of Athens, 2009.

[18] Y. Liu, K. Ramani, Robust principal axes determination for point-based shapes using least median of squares, Computer Aided Design 41 (2009) 293–305.

[19] D.V. Vranic, D. Saupe, 3D model retrieval, in: Spring Conference on Computer Graphics and its Applications, 2000.

[20] K. Fischer, B. Gartner, The smallest enclosing ball of balls: combinatorial structure and algorithms, International Journal of Computational Geometry and Applications 14 (4–5) (2004) 341–378.

[21] T. Napoleon, H. Sahbi, From 2d silhouettes to 3d object retrieval: contributions and benchmarking, EURASIP Journal on Image and Video Processing (2010) 17.

[22] M. Fischler, R. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Communications of the ACM 24 (1981) 381–395.

[23] Y. Chen, X. S. Zhou, T. Huand, One-class svm for learning in image retrieval, in: Proceedings of the International Conference on Image Processing, 2001.

[24] M. Manevitz, Y. Malik, One-class svms for document classification, Journal of Machine Learning Research 2 (2002) 139–154.

[25] P. Papadakis, I. Pratikakis, T. Trafalis, T. Theoharis, S. Perantonis, Relevance feedback in content-based 3d object retrieval: a comparative study, Computer-Aided Design and Applications 5 (5) (2008) 753–763.

[26] S. Tzikopoulos, M. Mavroforakis, H. Georgiou, N. Dimitropoulos, S. Theodoridis, A fully automated scheme for mammographic segmentation and classification based on breast density and asymmetry, Computer Methods and Programs in Biomedicine 102 (2011) 47–63.
[27] L. Wang, G. Xu, L. Guo, X. Liu, S. Yang, 3D reconstruction of head MRI based on one class support vector machine with immune algorithm, in: Proceedings of the IEEE International Conference on EMBS, 2007.
[28] Y. Zhang, N. Meratnia, P. Havinga, Adaptive and online one-class support vector machine-based outlier detection techniques for wireless sensor networks, in: Proceedings of the IEEE International Conference on Advanced Information Networking and Applications Workshops, 2009, pp. 990–995.
[29] S. Taeshik, M. Jongsub, A hybrid machine learning approach to network anomaly detection, Information Sciences 177 (18) (2007) 3799–3821.
[30] B. Scholkopf, J. Giesen, S. Spalinger, Kernel methods for implicit surface modeling, in: Advances in Neural Information Processing Systems, 2005, pp. 1193–1200.
[31] C.-C. Chang, C.-J. Lin, LIBSVM: a library for support vector machines, software available at: ⟨http://www.csie.ntu.edu.tw/~cjlin/libsvm⟩, 2001.
[32] F. Rong-En, C. Pai-Hsuen, L. Chih-Jen, Working set selection using second order information for training support vector machines, Journal of Machine Learning Research 6 (2005) 1889–1918.
[33] D. Tax, R. Duin, Support vector domain description, Pattern Recognition Letters 20 (11–13) (1999) 1191–1199.
[34] B. Scholkopf, A. Smola, Learning with Kernels, The MIT Press, 2002.
[35] P. Shilane, P. Min, M. Kazhdan, T. Funkhouser, The princeton shape benchmark., in: Shape Modeling International, 2004, pp. 167–178.
[36] R. Veltkamp, F. ter Haar, Shrec2007 3d shape retrieval contest, in: Technical Report, Department of Information and Computing Sciences, Utrecht University, 2007.
[37] R. Fang, A. Godill, X. Li, A. Wagan, A new shape benchmark for 3D object retrieval, in: International Symposium on Advances in Visual Computing, 2008.
[38] P. Papadakis, I. Pratikakis, T. Theoharis, S. Perantonis, Panorama: a 3d shape descriptor based on panoramic views for unsupervised 3d object retrieval, International Journal of Computer Vision 89 (2010) 177–192.
[39] A. Bronstein, M. Bronstein, R. Kimmel, Efficient computation of isometry-invariant distances between surfaces, SIAM Journal of Scientific Computing 28 (2006) 1812–1836.
[40] Z. Lian, A. Godil, T. Fabry, T. Furuya, J. Hermans, R. Ohbuchi, C. Shu, D. Smeets, P. Suetens, D. Vandermeulen, S. Wuhrer, Shrec'10 track: non-rigid 3d shape retrieval., in: Proceedings of the Eurographics/ACM SIGGRAPH Symposium on 3D Object Retrieval, 2010.
[41] Z. Lian, A. Godil, B. Bustos, M. Daoudi, J. Hermans, S. Kawamura, Y. Kurita, G. Lavoue, H. Nguyen, R. Ohbuchi, Y. Ohkita, Y. Ohishi, F. Porikli, M. Reuter, I. Sipiran, D. Smeets, P. Suetens, H. Tabia, D. Vandermeulen, Shrec'11 track: Shape retrieval on non-rigid 3d watertight meshes, in: Proceedings of the Eurographics/ACM SIGGRAPH Symposium on 3D Object Retrieval, 2011.

Panagiotis Papadakis received the University degree in Informatics and Telecommunications in 2005 and the D. Phil. degree in Information Technology from the National University of Athens, Greece, in 2009. From 2010 until 2012 he had been a post-doctoral researcher at ALCOR Cognitive Robotics Laboratory of DIAG, University of Rome "La Sapienza" in Italy, pursuing research in the field of Urban Search and Rescue Robotics. He is currently a post-doctoral researcher at INRIA, Equipe e-Motion, at Grenoble, France. His research interests span the fields of computer vision, machine learning and robotics with a special focus on robot perception, shape analysis & matching and robot motion planning.