

6DoF object pose measurement by a monocular manifold-based pattern recognition technique

Rigas Kouskouridas, Konstantinos Charalampous
and Antonios Gasteratos

Laboratory of Robotics and Automation, Department of Production and Management Engineering,
Democritus University of Thrace, Building I, Vasilissis Sophias 12, GR-671 00 Xanthi, Greece

E-mail: rkouskou@pme.duth.gr, kchara@pme.duth.gr and agaster@pme.duth.gr

Received 11 May 2012, in final form 28 August 2012

Published 17 October 2012

Online at stacks.iop.org/MST/23/114005

Abstract

In this paper, a novel solution to the compound problem of object recognition and 3D pose estimation is presented. An accurate measurement of the geometrical configuration of a recognized target, relative to a known coordinate system, is of fundamental importance and constitutes a prerequisite for several applications such as robot grasping or obstacle avoidance. The proposed method lays its foundations on the following assumptions: (a) the same object captured under varying viewpoints and perspectives represents data that could be projected onto a well-established and highly distinguishable subspace; (b) totally different objects observed under the same viewpoints and perspectives share identical 3D pose that can be sufficiently modeled to produce a generalized model. Toward this end, we propose an advanced architecture that allows both recognizing patterns and providing efficient solution for 6DoF pose estimation. We employ a manifold modeling architecture that is grounded on a part-based representation of an object, which in turn, is accomplished via an unsupervised clustering of the extracted visual cues. The main contributions of the proposed framework are: (a) the proposed part-based architecture requires minimum supervision, compared to other contemporary solutions, whilst extracting new features encapsulating both appearance and geometrical attributes of the objects; (b) contrary to related projects that extract high-dimensional data, thus, increasing the complexity of the system, the proposed manifold modeling approach makes use of low dimensionality input vectors; (c) the formulation of a novel input–output space mapping that outperforms the existing dimensionality reduction schemes. Experimental results justify our theoretical claims and demonstrate the superiority of our method comparing to other related contemporary projects.

Keywords: pose estimation, object recognition, manifold learning, mutual information

(Some figures may appear in colour only in the online journal)

1. Introduction

Recognizing patterns in measurement science is of fundamental importance since it enables the efficient accomplishment of several vital tasks in a plethora of diverse applications. The bidirectional connection between pattern recognition techniques and measurement technology was recognized long ago [1]. On the one hand, measurements need

to be classified so that the measured signal can be further studied; on the other hand, any pattern recognition based technique is of no use unless real data are gathered by a measuring system. Pattern recognition is the scientific field which aims to provide algorithms able to categorize objects into classes. The type of object/pattern is application specific including any type of signal, such as sound waveforms, images and measurements. Pattern recognition is a key element in

decision making systems. Particularly, in imaging systems where a camera records images, a pattern classification method should be employed to describe, analyse and interpret the sequential frames [2, 3].

In the field of computer vision and image understanding, extensive emphasis is given to the 3D object pose estimation problem, due to the fact that there is still no general ground solution entailing both sufficient performance and high generalization capacities. It is palpable that an accurate measurement of the orientation of a target relative to the camera's plane is of great importance for several basic computer vision tasks [4–6]. According to the literature, the techniques dedicated to the 3D pose estimation problem could be categorized into two major research streams. The first category includes methods that utilize numerous model images of training objects to construct large databases, which are established in a supervised manner [7, 8]. The second category, part-based (or constellation-based) architectures entail unsupervised learning capabilities, whilst aiming at extracting highly distinguishable areas of an object with a view to empower generalization abilities [9, 10]. However, an advanced imaging technique that mimics the remarkable skills of humans, estimating the relative pose of objects given an initial hypothesis, has yet to be built.

In this paper, we propose a sophisticated framework capable of both recognizing objects and estimating their pose in the 3D working space. Following the intuition that (a) one object viewed under varying perspectives lays on a well-projected subspace (figure 1(right)) and (b) different objects captured under similar viewpoints share identical poses (figure 1(left)), we formulate a manifold modeling process that depends on the attributes of mutual information and a constellation-based structure, respectively. The latter entails the unsupervised clustering of the extracted visual cues in order to find the corresponding centers that hold both appearance and geometrical data. We then employ two individual modules for recognition and pose estimation purposes, respectively. Regarding the recognition module, we propose a modified approach of a known dimensionality reduction technique that constructs a similarity matrix based on mutual information [11, 12] among objects and then seek a low-dimensional representation that preserves the local structure of the objects in the initial high-dimensional space. Once the data are projected into the sub-space a support vector machine (SVM) [13] classifier is responsible for providing accurate recognition. Concerning the 3D pose estimation module, the distances of each of the computed clusters from a given one are taken into account through the establishment of the manifolds. Although, the latter are of low dimensionality, they are capable of remarkably distinguishing similar poses of different classes. Finally, an accurate measurement of the 3D orientation of a testing object is obtained through a neural network-based solution that incorporates a novel input–output space mapping. The main contributions of our paper are: (i) we employ a constellation-based architecture that, compared to other related works [9, 10], bypasses the part selection process using unsupervised clustering, while extracting local patches encapsulating both appearance and geometrical attributes of

the objects; (ii) we propose a new manifold modeling approach that, opposed to [14, 15], establishes input vectors of lower dimensionality, whilst utilizing numerous datasets without restricting the 3D pose estimation to cars only; (iii) both the recognition and the 3D pose estimation modules are based only on the properties of the established manifolds, thus, reducing the complexity of the proposed system; (iv) we formulate the new input–output space mapping that outperforms the conventional dimensionality reduction techniques widely used in image understanding applications. Experimental results prove that there are indeed compact manifolds that can be efficiently learnt from highly representative datasets and that enable accurate object recognition and 3D pose estimation. Finally, we comparatively evaluated the performance of our method with other related projects, whilst experimental results provide evidence of low generalization error.

The remainder of this paper is structured as follows. In section 2, we review existing approaches that delve into the recognition, 3D pose estimation and manifold modeling problems. The overview of the proposed method is presented in section 3, while the manifold modeling approach that incorporates the recognition and 3D pose estimation modules is introduced in section 4. Analytical experimental evaluation along with comparative results are available in section 5. Last, conclusions are drawn in section 6 along with some final notes and an outlook to future work.

2. Related work

In order to adequately carry out 3D object pose estimation tasks, computer vision methods should tackle one important cascading issue that hinders their effective appliance. The 2D–3D image point correspondence problem is of fundamental importance since its sufficient solution directly enables the clarification of the 3D object pose estimation task. Throughout the years, scholars have primarily focused on designing and implementing techniques that solve the 3D pose problem by applying conventional mathematical models, e.g. planar homographies [16], contour extraction [17] or line-based estimations [18]. However, deterministic approaches are adopted to solve particular problems and, therefore, they lack generalization capabilities in totally unknown tasks. Toward this end, contemporary research efforts aim at designing sophisticated algorithms that incorporate machine learning subroutines [19, 20]. As a result, extensive emphasis is given to the process of building the respective input vectors along with their internal structure and dimensionality. In [21], the latter is reduced through PCA, whilst an accurate estimation of the 3D pose of an object is obtained via an appearance-based method. Additionally in the same work, the proposed neural network is trained with the resilient backpropagation method, while only two DoFs are sufficiently computed. To address the high dimensionality issue, Yuan *et al* [22] proposed the usage of nonlinear principal component analysis (PCA) for the extraction of the input vectors. Notwithstanding their sufficient classification results, both methods [21, 22] sort interpolation capacities mainly due to their oversampled datasets resulting from a sampling interval of 3°.

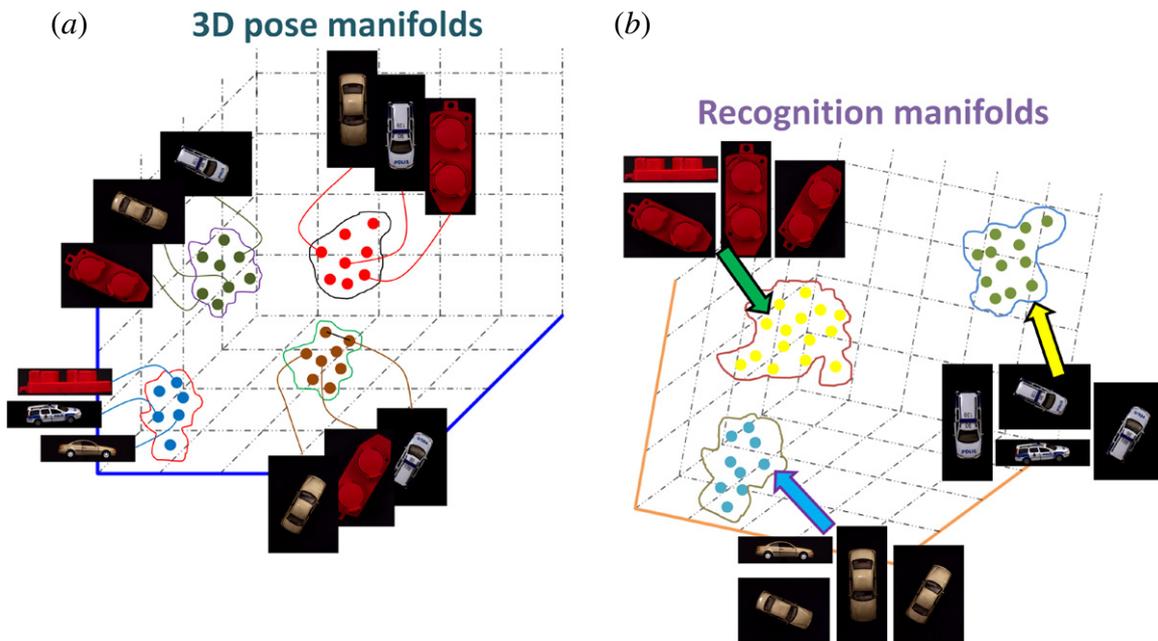


Figure 1. This figure illustrates the key idea underlying the proposed framework. Following the intuition that (a) different objects viewed under the same geometrical positions share identical poses and (b) different views of the same object ‘lay on’ the same subspace, we model the 3D pose and recognition manifolds, respectively. Both established manifolds are of low dimensionality and are capable of remarkably distinguishing 3D pose and recognition patterns, respectively.

In this paper, we aim at computing the full 3D pose of an object by proposing an approach that, among others, encompasses a constellation-based (or part-based) architecture. Generally, image understanding techniques of this field focus on learning highly discriminative object pose models by extracting individual parts of targets that exhibit apparent local information [23, 24]. In these works, experimental results prove that the efficiency of the 2D–3D point correspondence subroutine is directly related to the process of picking the most prominent visual patterns available. More recently, in [25] a method that emphasizes revealing the liaison between the built input vectors and the corresponding performance measurement was presented. Furthermore, as was shown in [9], any part-based architecture is characterized by its remarkable ability to efficiently portray object models by simply linking together diagnostic regions of objects from different viewpoints.

Many dimensionality reduction (DR) techniques have been proposed in the past few decades. In general, they can be separated into supervised and unsupervised ones, i.e. the label of each data sample $\mathbf{x} \in \mathbb{R}^m$, where m is the dimension of the sample, is considered to be known during the execution of the corresponding procedure. The most representative algorithms of those major categories are PCA [26] and linear discriminant analysis (LDA) [27]. PCA is a linear unsupervised technique which, given a data matrix $X_{m \times n}$, where n is the number of samples, seeks to project the data to an orthogonal sub-space such that the variance of the data is maximized. A well-known linear supervised technique is LDA, which is a generalization of Fisher discriminant analysis (FDA) that is suitable for multi-class problems. LDA searches for a sub-space to project the data such that the ratio of between-class to within-class scatter is maximized.

There are many other linear methods, such as independent component analysis (ICA) [28], where the resulting projections are not necessarily orthogonal, yet their statistical independence is maximized. Regarding the nonlinear techniques, a considerable portion of them are kernelized versions of linear algorithms. After the ‘kernel trick’ [13] was introduced via support vector theory [13], it was widely applied in many linear dimensionality reduction methods, such as kernel-PCA [29] and generalized discriminant analysis [30].

Other newly developed nonlinear techniques are Laplacian eigenmap [31], local linear embedding (LLE) [32] and ISOMAP [33]. Recently, a general framework has been proposed [34], named graph embedding and providing a theory that unifies most of the dimensionality reduction techniques, such as the aforementioned. Along with the data matrix $X_{m \times n}$ and a similarity matrix $F_{n \times n}$, graph embedding scheme constructs a graph $G = \{X, F\}$. Weight matrix F encapsulates geometrical or statistical properties among data samples in the original input space. The purpose is to compute a low-dimensional manifold that preserves the similarity characteristics among data, as documented in F . It is worth mentioning that the graph embedding scheme supports, besides the linear approach, also a kernelized and a tensorized expansion.

Regarding the object recognition module, it is based on the local preserving projections (LPP) [35] dimensionality reduction technique, which has been successfully applied in face recognition problems. LPP computes a similarity matrix utilizing the L_2 norm, while the proposed framework exploits the mutual information (MI) criterion among two images. Moreover, the SVM classifier creates a separation hyperplane based on the geometrical characteristics of the

training dataset, which is an intriguing property, suitable for the discrimination of low-dimensional manifolds. Concerning the processes of building a constellation-based architecture and manifold modeling, the closest works to our paper are presented in [10, 14, 15]. The ‘natural 3D markers’, as they appear in [10], comprise a part-based structure that extracts four or five close feature points, which encapsulate distinctive photometric properties and possess equal distribution over the visible surface of the testing objects. Notwithstanding the reported efficacy of the method, as comparative experimental results prove, ‘natural 3D markers’ fail to construct compact and abstract representations of the 3D objects, whilst being less tolerant to partial occlusions. Regarding the proposed manifold modeling module, our work is influenced by those presented in [14, 15]. In both papers, the authors showed that part-based forms, opposed to holistic-based ones, are capable of controlling more efficiently casual imaging disturbances, i.e. background clutter, partial occlusions. Moreover, they proved that solid pose and recognition manifolds can be adequately learnt from several databases. However, our work outperforms these methods due to the fact that (a) it requires less supervision during the learning procedure without restricting the 3D pose estimation to cars only; (b) the resulting manifolds are of very low dimensionality, unlike in [14] and [15], which influences directly the performance of any regressor or classifier.

3. Overview of our approach

The key idea underlying the proposed method, as already depicted in figure 1, represents an advanced dimensionality handling process that aims at establishing novel manifolds for simultaneous recognition of unknown objects and the estimation of their 3D pose. With a view to the reader’s better understanding, in figure 2, we illustrate the main building blocks of our method. Initially, labeled datasets, for both recognition and 3D pose purposes, are divided into training and testing subsets, respectively. We then employ the manifold modeling module that aims at inducing an architecture capable of (a) classifying numerous testing objects into the corresponding categories and (b) projecting similar poses of different objects onto the same subspace. Recognition manifolds impose upon the context of mutual information framework that is responsible for the efficient design of the adjacency matrix needed. The latter is given as input to an SVM-based classifier, which is trained over a large dataset of training examples and that is responsible for providing recognition results for new testing patterns supplied to the system.

Regarding the 3D pose manifolds, unlike in [15], where ‘alignment’ and ‘expansion’ operations are employed, in order to suffice for proper discrimination of high-dimensional vectors, we propose a manifold modeling routine that encompasses distance handling operations over the extracted centers computed through unsupervised clustering. As a follow-up step, the resulting manifolds are used to train a radial basis functions (RBF)-based regressor that entails a novel input–output space mapping. Finally, accurate measurements of the 3D pose of testing objects is obtained by simulating the

proposed regressor. Opposed to [15], our method is capable of establishing highly distinguishable manifolds without requiring the testing objects to belong to the training set and the same object category (e.g. cars).

4. Manifold modeling

4.1. Recognition manifolds

This section presents the object recognition module used in the proposed approach. The aforementioned scheme is a two step process, first a modified feature extraction technique is applied and the resulting low-dimensional data constitute the input to the SVM classifier [13]. MI [36] is used as a similarity measure between two images [12, 37], yielding remarkable results. MI has been successfully employed as a similarity measure, specifically in the medical imaging domain [12], primarily due to its robustness to outliers and its minimum complexity. In particular, it measures the statistical dependence between two random variables, which in our case correspond to the context of two images or the amount of information contained in an image compared to a baseline one. Moreover, it can be equivalently interpreted as a measure that defines the degree of dependence between a pair of images. This measure traces its roots to information theory and is defined as follows:

$$I(R_{v1}, R_{v2}) = \sum_{r_{v1} \in R_{v1}} \sum_{r_{v2} \in R_{v2}} P(r_{v1}, r_{v2}) \log \frac{P_{R_{v1}, R_{v2}}(r_{v1}, r_{v2})}{P_{R_{v1}}(r_{v1})P_{R_{v2}}(r_{v2})}, \quad (1)$$

where R_{v1}, R_{v2} are two random variables, $P_{R_{v1}}(r_{v1})$ and $P_{R_{v2}}(r_{v2})$ are the marginal probability distribution functions of R_{v1} and R_{v2} , respectively, $P_{R_{v1}, R_{v2}}(r_{v1}, r_{v2})$ indicates the joint probability function and r_{v1}, r_{v2} are variables that run through the sets of all possible values of R_{v1}, R_{v2} .

MI’s relation to entropy is expressed via

$$I(R_{v1}, R_{v2}) = H(R_{v1}) - H(R_{v1}|R_{v2}) \quad (2)$$

$$= H(R_{v2}) - H(R_{v2}|R_{v1}) \quad (3)$$

$$= H(R_{v1}) + H(R_{v2}) - H(R_{v1}, R_{v2}) \quad (4)$$

$$= H(R_{v1}, R_{v2}) - H(R_{v1}|R_{v2}) - H(R_{v2}|R_{v1}), \quad (5)$$

where $H(R_{v1})$ and $H(R_{v2})$ are the entropies of R_{v1}, R_{v2} , $H(R_{v1}|R_{v2})$ and $H(R_{v2}|R_{v1})$ are the conditional entropies of R_{v1} given R_{v2} and R_{v2} given R_{v1} , respectively, and $H(R_{v1}, R_{v2})$ is their joint entropy [38]. In this paper, MI is employed as a distance measure in a similarity matrix; thus, the normalized variation of information [39] metric is used. Normalized variation of information is defined as

$$d_{VI} = H(R_{v1}|R_{v2}) + H(R_{v2}|R_{v1}) \quad (6)$$

$$M_{VI} = \frac{d_{VI}}{H(R_{v1}, R_{v2})} \quad (7)$$

and the inequality $M_{VI} \leq 1$ holds always as proven in [39].

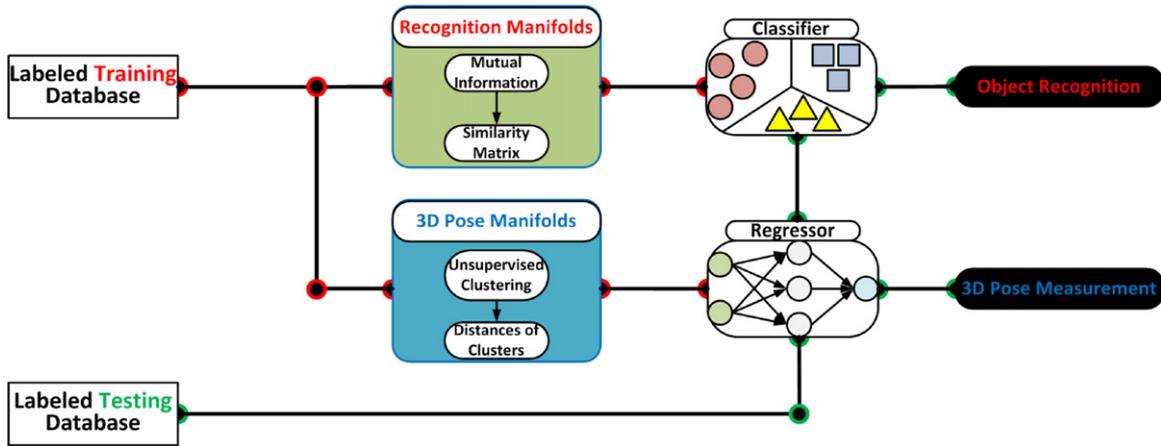


Figure 2. The first stage of the proposed method incorporates the division of the available datasets into the corresponding testing and training subsets. The latter are provided as input to the process of manifold modeling that establishes the recognition and 3D pose manifolds. As a follow-up step, we train an SVM-based classifier and an RBF-based regressor in order to obtain reliable recognition results and to acquire accurate 3D pose measurements, respectively. Finally, we evaluate the performance of our method in cases where an unknown testing particle is provided to the system.

Locality preserving projections (LPP) [35] is a linear dimensionality technique that comes under the graph embedding framework and it is considered to be a linear approximation of Laplacian eigenmap [31], a nonlinear technique as mentioned in section 2. Unlike the PCA, which aims to preserve the global structure of the data, LPP targets to preserve the local structure. The proposed modified LPP scheme defines the similarity via the normalized variation of information metric. Initially, a graph G is constructed having n nodes. For each pair (i, j) of nodes, where $i, j = 1, 2, \dots, n$, an edge is drawn if the data sample \mathbf{x}_i lays among the k nearest neighbors (k -NN) of \mathbf{x}_j , where $k \in \mathbf{N}$. The similarity matrix $F_{n \times n}$ is defined as follows:

$$F_{ij} = \begin{cases} \exp(-(1 - M_{VI})), & G_{ij} = 1 \\ 0, & G_{ij} = 0. \end{cases}$$

That is, $G_{ij} = 1$ indicates the existence of an edge whereas $G_{ij} = 0$ the absence. The following minimization criterion aims to preserve the similarity of the initial input space to the corresponding low-dimensional space [35]:

$$\sum_{i=1}^n \sum_{j=1}^n (y_i - y_j)^2 F_{ij} \quad (8)$$

By applying algebra formulations, the minimization criterion equals

$$\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (y_i - y_j)^2 F_{ij} = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\mathbf{w}^T \mathbf{x}_i - \mathbf{w}^T \mathbf{x}_j)^2 F_{ij} \quad (9)$$

$$= \mathbf{w}^T X(D - F)X^T \mathbf{w} \quad (10)$$

$$= \mathbf{w}^T XLX^T \mathbf{w}, \quad (11)$$

where D is a diagonal matrix and its elements are the corresponding column summations of F , i.e. $D_{ii} = \sum_j F_{ij}$. $L = D - F$ is the Laplacian matrix. A constraint is also

introduced: $\mathbf{w}^T XDX^T \mathbf{w} = 1$ and thus the minimization problem results in

$$\min_{\mathbf{w}} \frac{\mathbf{w}^T XLX^T \mathbf{w}}{\mathbf{w}^T XDX^T \mathbf{w}}. \quad (12)$$

The minimization criterion yields the optimal projection vector \mathbf{w} by solving the following generalized eigenvalue problem:

$$\mathbf{w}^T XLX^T \mathbf{w} = \lambda \mathbf{w}^T XDX^T \mathbf{w}. \quad (13)$$

According to the above analysis, our modified LPP method and the original one presented in [35] form the graph G via the k nearest neighbors method, which through the employment of the Euclidean distance as a proximity measure extracts geometrical attributes of the sought object. In addition, the original LPP methodology forms the weight matrix F via the application of a heat kernel, which is also based on the Euclidean distance between the corresponding samples. In this paper we have substituted the heat kernel by the MI, which allows us to handle complex relationships between the intensities in a pair of images with altering viewpoints. Thus, the resulting matrix F is significantly sparser than the original one, while in most of the cases, samples from different classes produce a zero value in F . Sparsity is a key element in order to encode knowledge, therefore an essential tool in order to achieve great generalization capabilities. Moreover, the proposed framework allows the exploitation of the relationship among corresponding pixels contrasted to the original LPP method that extracts only geometrical and neighborhood attributes of the objects.

4.2. 3D pose manifolds

During the pose manifold estimation phase, we aim at building a part-based architecture that depends on the extracted visual cues over the surface of the object. In order for this structure to be robust and feasible to be constructed, its members should represent visual attributes of the objects that are not only distinguishable enough, but also hold a significant quantity of data. Toward this end, we address this issue by employing

a part-based scheme that extracts key points, over the surface of an object, which enjoy appearance-based properties along with geometrical ones, i.e. part's location distribution. Despite the fact that the same notion characterizes the work presented in [15], it fails to generalize to unknown objects, restricting their performance to trained ones. We believe this is due to the fact that the procedure of part extraction craves extensive supervision since each part corresponds to a realization of the probability density function (PDF) associated with the joint distribution of appearance and geometry attributes of an object. In order to enhance the generalization capacities of our method, we revise unsupervised clustering capabilities with a view to extract representative centers over the abstracted features of the object. Concretely, we account for appearance-based data by employing the SIFT descriptor [40], without being limited to the particular selection. The extraction of SIFT features is accompanied by a homography-based RANSAC [41] for outlier removal. Additionally, geometrical attributes of the objects are encapsulated through the implementation of the K -means unsupervised clustering method.

Let $Q(T, o_i|a_j)$ representing the raw intensity values for the three-channeled image I of object $o_i \in [o_1, o_2, \dots, o_m]$ with pose $a_j \in [a_1, a_2, \dots, a_n]$. We then extract ρ key points of interest and denote them as $I(\mathbf{u}^\rho, o_i|a_j)^d$, $\mathbf{u} \in \mathbb{R}^2$, by integrating the SIFT detector and descriptor, respectively. Here, d corresponds to the dimensionality of the descriptor, which is chosen to be 128, whilst the resulting vector is normalized to unit length, in order to maintain invariance to affine changes in illumination. Then we define the following notions:

- $\mathbf{u} \in \mathbb{R}^2$ are the locations of the extracted appearance-based features,
- $c^{(\rho)}$ represent the index of cluster $(1, 2, \dots, l)$ to which feature ρ at location \mathbf{u}^ρ is randomly assigned,
- μ_l correspond to the position of cluster centroid l ($\mu_l \in \mathbb{R}^2$), and
- $\mu_{c^{(\rho)}}$ represent the centroid of cluster to which example \mathbf{u}^ρ is assigned after one step of the algorithm.

The goal is to find those clusters that minimize the following objective function:

$$J(c^{(1)}, \dots, c^{(\rho)}, \mu^{(1)}, \dots, \mu^{(l)}) = \frac{1}{\rho} \sum_{i=1}^{\rho} \|\mathbf{u}^i - \mu_{c^{(i)}}\|^2. \quad (14)$$

This is an iterative process that executes till the convergence of the cost function $J(c^\rho, \mu^l)$. The computed clusters are considered as subsets denoted as $\xi = \langle \mu_l, o_i|a_j \rangle$ and represent the built part-based structure for object o_i observed under pose a_j .

For the purposes of the 3D pose manifold establishment and regarding the $(o_i^*|a_j^*)$ training object, we assume that the computed graphs appoint the feature vector \mathbf{e} with the set of all feature vectors being $\mathcal{E} = \{\mathbf{e}_c : c \text{ is the number of clusters organized as vectors}\}$. Additionally, let $\mathbf{e}^* \in \mathcal{E}^*$ be a randomly selected example vector drawn from $\mathcal{E}^* \subseteq \mathcal{E}$. The proposed manifold modeling frame-

work proceeds by computing the L_2 norm between the vector $\mathbf{e}_i \in \mathcal{E}$ and the anchor point \mathbf{e}^* :

$$\mathbf{v}^i = \|\mathbf{e}_i - \mathbf{e}^*\|^2 = \sum_{i=1}^c \{\mathbf{e}_i - \mathbf{e}^*\}^2. \quad (15)$$

As a next step, we learn a regressor that provides a mapping from a set of input variables $\mathbf{v} = [v_1, \dots, v_d]$, belonging to a feature-space \mathcal{V} , to a modeled output variable $\omega = \omega(\mathbf{v}; \zeta) \in \Omega$, with ζ denoting the vector of the adjustable parameters. The ultimate goal of our system is to learn a regressor $\tau : \mathcal{V} \rightarrow \Omega$ from an *a priori* training dataset $\{\mathbf{v}^n, \omega^n\}$ in order to efficiently approximate the output Ω_t , when an unknown example \mathcal{V}_t is provided. The proposed architecture encompasses a new input–output mapping procedure that does not require common dimensionality reduction operations. In particular, opposed to [15, 21] where the input vectors are of very high dimensionality, the established manifolds are of low dimensionality and are directly fed into the regressor. The most common approach for the input normalization is the linear transformation of given vectors so that input variables are independent. Such information transformation is generally based on the mean removal method and results in sets of input vectors with zero mean and unit standard deviation. However, this linear rescaling treats input variables as independent, while in most cases they are not [42]. With a view to achieve an efficient solution to this problem, we adopted a more prominent strategy which allows correlations amongst variables. Therefore, the input variable \mathbf{v}^i is organized as a vector $\mathbf{v} = [v_1, \dots, v_c]^T$, while the sample mean vector and the covariance matrix with respect to the β data points of the training set are

$$\bar{\mathbf{v}} = \frac{1}{\beta} \sum_{n=1}^{\beta} \mathbf{v}^n$$

$$\Sigma = \frac{1}{\beta - 1} \sum_{n=1}^{\beta} (\mathbf{v}^n - \bar{\mathbf{v}})(\mathbf{v}^n - \bar{\mathbf{v}})^T. \quad (16)$$

This normalization results in vectors with the input variables given by the following formula:

$$\tilde{\mathbf{v}}^n = \Lambda^{-1/2} B^T (\mathbf{v}^n - \bar{\mathbf{v}}), \quad (17)$$

where $B = (\mathbf{b}_1, \dots, \mathbf{b}_c)$ and $\Lambda = (\lambda_1, \dots, \lambda_c)$ correspond to the eigenvectors and eigenvalues, respectively, which are calculated from the covariance matrix $\Sigma \mathbf{b}_\phi = \lambda_\phi \mathbf{b}_\phi$.

5. Experimental results

In order to evaluate the performance of the proposed method and, essentially, the discrimination ability of the established manifolds, we utilize numerous databases containing several objects under varying viewpoints. The aforementioned modules were tested utilizing the ETH-80 dataset. ETH-80 consists of eight different classes; each class containing ten different instances (for example, class car varies in color, shape, convertible, etc) and each instant has been captured at 41 different orientations. The image size of all images in the dataset is 128×128 pixels. Moreover, in order to evaluate the

experimental results and the generalization capabilities of the corresponding modules, K-fold cross validation was applied [43]. Regarding the databases utilized for the establishment of the 3D object pose manifolds, we would like to highlight the severe lack of datasets dedicated to 3D pose evaluation. To address this issue, we utilize a rather expanded labeled training set that comprises (a) images of real objects belonging to the COIL-100 database [44] and (b) a large collection of synthetically rendered objects, which are shot every 5° and are available on-line [45]. Additionally, the labeled testing set contains images of the CVL database [46] that includes several objects captured under varying viewpoints in both cluttered and uncluttered environments. Finally, we would like to state that the proposed 3D pose module can be easily adjusted to include any feature extraction or clustering algorithm and it is not restricted by the selection of SIFT and K -means, respectively. Indeed, several mixtures of detectors/descriptors and clustering techniques were assessed and the proposed one excels in accuracy and robustness. Yet it should be mentioned that if execution speed is the criterion, a new set of SURF-based features, as proposed in [47], should be employed.

5.1. Recognition manifolds

This section provides the evaluation of the proposed recognition scheme. This work is experimentally compared with state-of-the-art object recognition schemes that incorporate well-known dimensionality reduction techniques together with strong classifiers. More specifically, the PCA+SVM that the method utilized for cephalometric analysis [48] exhibit remarkable accuracy. In addition, GDA+SVM [49] was also used for solving the classification problem of detecting the existence of lung cancer. Recently, the authors in [50] combined the LPP dimensionality reduction technique with the SVM classifier forming a robust categorization method that exhibits remarkable accuracy in the language recognition problem. These methods constitute the state of the art in the field of dimensionality reduction aiming to solve efficiently different pattern recognition problems. As we have previously mentioned in section 4.1, the goal of the proposed method is to develop an enhanced methodology based on the initial LPP algorithm, suitable for object recognition tasks. Therefore, it is of great importance to comparatively examine the performance of the proposed methodology against the state-of-the-art ones.

Due to the fact that SVM is a binary classifier, the one-versus-all [51] technique was adopted. The experimental procedure involves the linear SVM. Different kernel types were exhaustively tested (i.e. polynomial, Gaussian, sigmoid and linear) during the training and testing procedure. However, the linear one exhibited higher classification rates opposed to other kernels, while such behavior of the SVM classifier can be justified through the acquisition of a maximum margin linear separation. As shown in [52], the derivation of a large margin on the training data set leads with high probability to a classifier endowed with large generalization capabilities. Additionally, our theoretical claims are experimentally justified since our method derives such a margin that enables the adequate

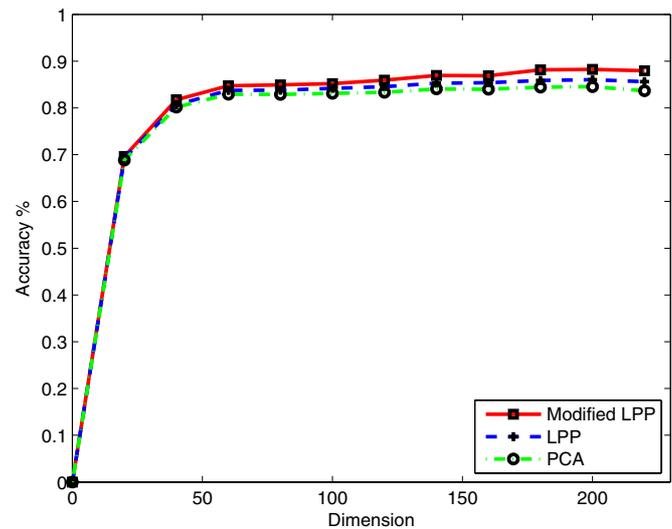


Figure 3. The impact of recognition manifold dimensionality concerning the classification results. The number of dimensions beyond 180 cause decreasing accuracy results.

linear separation of the testing samples when projected onto the new sub-space. Parameter selection holds a key role in overall performance; thus, in order to be maximized, several parametric values have to be determined. The parameters that have to be estimated are the number of neighbors in modified LPP and LPP and the number of dimensions in modified LPP, LPP and PCA, GDA technique was left out due to the fact that the number of dimensions is the number of classes minus one. Regarding the SVM, since the linear one was utilized, only the relaxation parameter C has to be tuned. Figure 3 depicts the efficiency of the aforementioned techniques with respect to the dimensionality of the recognition manifold. The classification accuracy was maximized when the module dimensionality was set equal to 180.

As explained in subsection 4.1, a graph needs to be constructed from both LPP and modified LPP module. In order for the graph to place the edges between nodes, the nearest neighbors of each sample are computed. Depending on the number of neighbors, the graph and, thus, the similarity matrix that is formed can be either dense or sparse. This parameter is also application dependent, the contribution of the number of neighbors is illustrated in figure 4. It is easily identified that the most suitable selection is the one of five neighbors.

Concerning the SVM classifier, the linear kernel was selected for the training procedure, and in order to maximize the resulting classification rate, the regularization parameter C had also to be fixed optimally. The performance of the corresponding module was maximized by setting parameter $C = 100$ as can be seen in figure 5. The one versus all methodology was utilized for the corresponding multi-class problem. The average recognition accuracy for each class separately together with their respective standard deviation is shown in table 1. These results debrief the per class average classification performance for each subordinate algorithm. In particular, the proposed method exhibits higher classification accuracy in seven out of eight classes, as compared with the benchmark methods. Additionally, by examining the standard

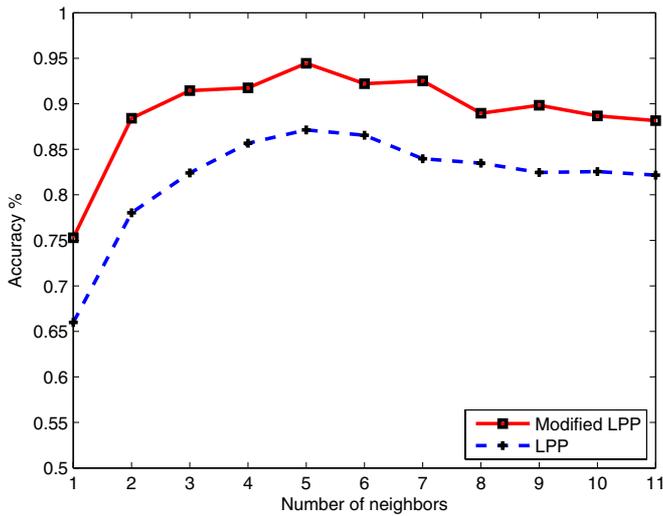


Figure 4. The effect of number of neighbors used in an intermediate step of both modified LLP and LPP techniques and the contribution of the corresponding parameter to classification accuracy.

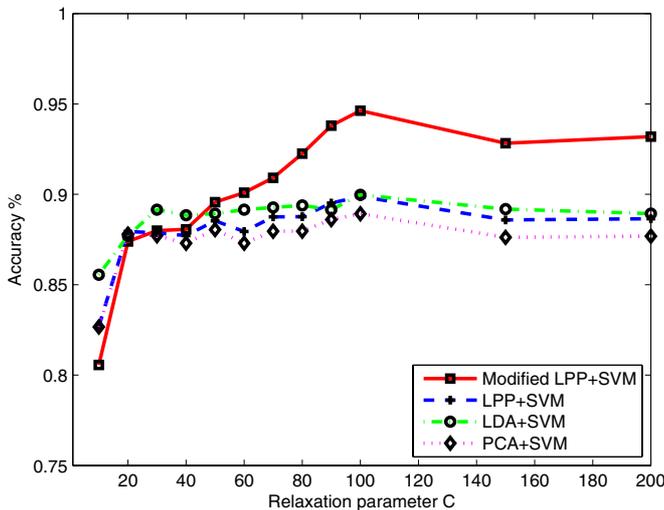


Figure 5. The influence of relaxation parameter C utilized by the SVM classifier on classification results.

deviation values, it is revealed that the modified LPP+SVM possesses more tight boundaries for the majority of the classes. The latter is an indication that the proposed dimensionality reduction method (modified LPP) exhibits higher performance rates than the initial one (LPP), thus boosting the generalization capabilities of the SVM classifier. An error bar diagram demonstrating the average overall accuracy of the methods under evaluation, plotted with two standard deviations is shown in figure 6. The fact that there is a slight overlap between the error bar of the proposed methodology and those of the benchmark methods indicates that the difference in categorization performance is statistically significant.

5.2. 3D pose manifolds

Building the training set. The training phase of the proposed module encapsulates the processes of building the training

set $\{\mathbf{v}^n, \omega^n\}$ and the training of the RBF-based regressor τ . The labeled training set comprises n training examples that correspond to images of o_j objects that are accompanied by the respective pose groundtruth measurements. Additionally, training examples that represent images of objects shot every 5° correspond to known object poses ω^n , which are efficiently measured via the process of placing each training target on a turntable. It is apparent that the construction of the training set $\{\mathbf{v}^n, \omega^n\}$ imposes upon an iterative procedure over n images of o_j objects. Furthermore, the task of finding matches between one image of an object and others depicting the same target under different viewpoints stands for a routine of paramount importance. We address this issue by introducing the tracking sensitivity parameter Δm , which (a) aims at minimizing the labor-intensive matching process by comparing one object's image to a limited neighbor images, i.e. if $\Delta m = 1$ the object's image is correlated only to photos depicting the same target with $+5^\circ$ and -5° pose discrepancies around any axis; (b) represents the range of the output of the regressor τ , i.e. if $\Delta m = 1 \rightarrow [-5^\circ, +5^\circ]$ or if $\Delta m = 6 \rightarrow [-30^\circ, +30^\circ]$; (c) given a moderate vision sensor capturing 30 frames per second, it offers a tracking ability of either 300 deg s^{-1} ($\Delta m = 1$) or 1800 deg s^{-1} ($\Delta m = 6$). The visual results of the proposed 3D pose module is illustrated in figure 7.

It is apparent that, during the training process of the proposed neural network-based solution, convergence is of great significance. Our work depends on the efficient establishment of highly discriminative manifolds that are applied on large datasets. Generally, network training with a large number of training examples constitutes an appealing convergence task, since, most of the time, the calculation of the local minimum instead of the global one is highly likely. In this work, we address this issue by adding noise to the input vectors during the training process, thus, expanding the training set by replicating random input vectors. Therefore, we are able to bootstrap the performance of the network avoiding local minima.

Comparison with related methods and dealing with occlusions.

The performance of the proposed 3D pose module was comparatively evaluated against other related works solving for the 3D pose problem. Particularly, we compare our method with (a) *neural network-based solution*: the work of Yuan *et al* [21] that embodies PCA for dimensionality reduction purposes; (b) *a part-based architecture*: the method of Hinterstoisser *et al* [10] for constellation-based pose estimation; (c) *a part-based architecture and manifold modeling method*: the related approach of Mei *et al* [15] for manifold modeling and 3D object pose estimation; (d) *a standard baseline technique*: the classic least-squares solution of Lowe [24].

The most common evaluation criterion of 3D pose measurements is the one that studies the performance of the respective work under partial occlusions, since the latter affect directly the efficacy of any computer vision application. Although humans are capable of simultaneously recovering the 3D pose of a target under difficult occlusion circumstances, such an inclination is limited in contemporary

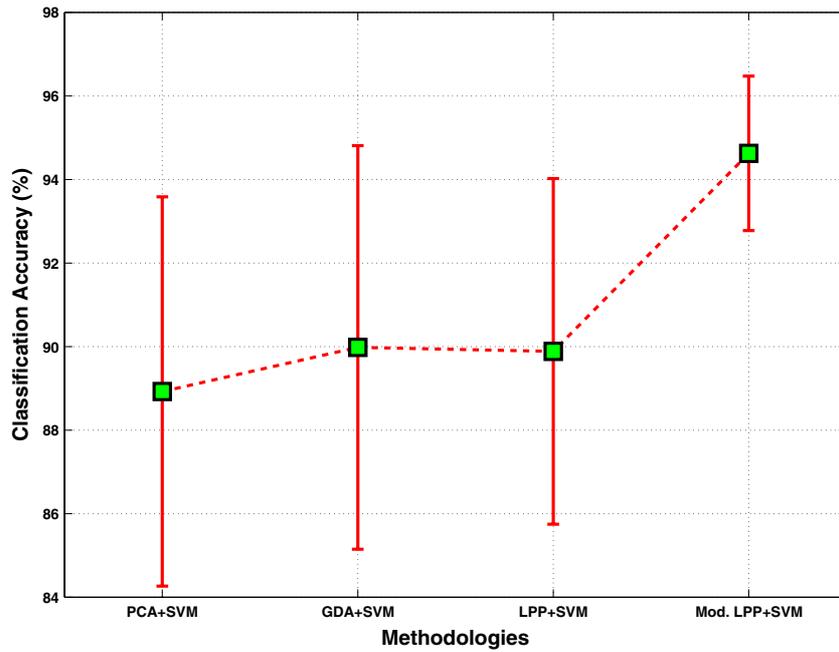


Figure 6. This figure illustrates the error bar diagram regarding the average overall accuracy of the methods under evaluation.

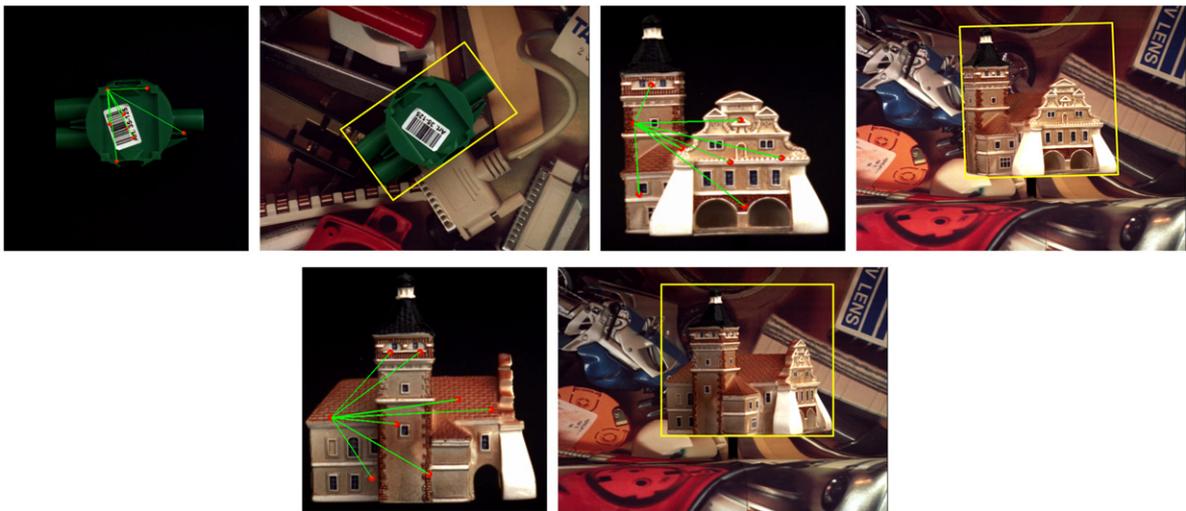


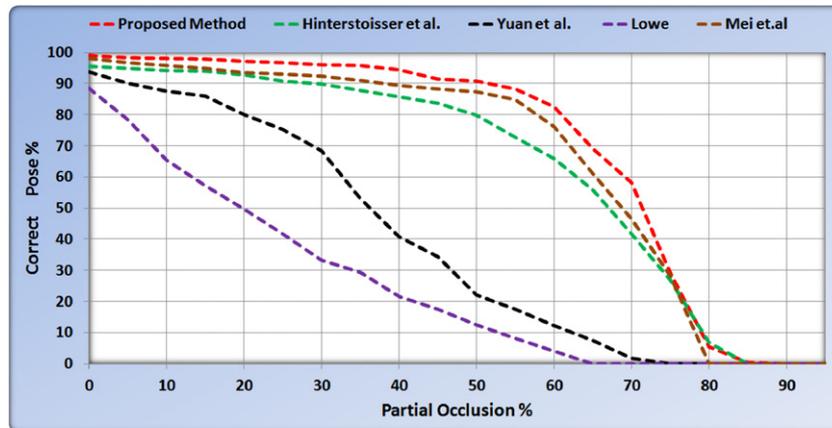
Figure 7. The visual outcome of the proposed 3D pose estimation framework for several test objects.

Table 1. Average classification accuracy (%) on the ETH-80 dataset.

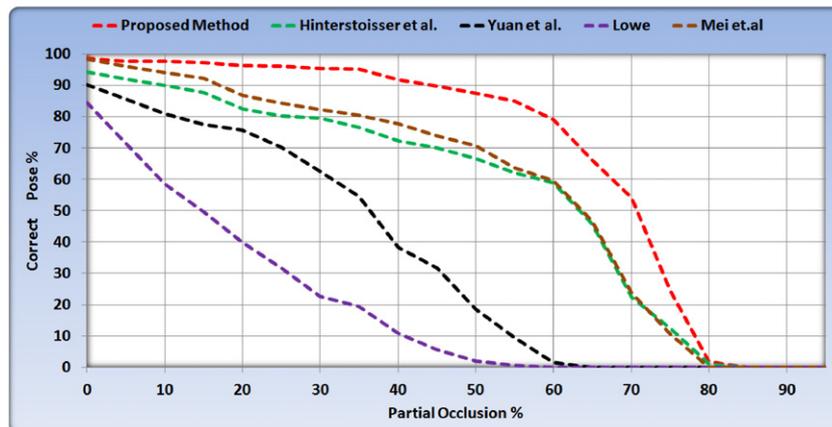
Class	PCA+SVM		GDA+SVM		LPP+SVM		Modified LPP+SVM	
	Accuracy	STD	Accuracy	STD	Accuracy	STD	Accuracy	STD
Apple	88.2146	4.2	89.2458	3.6	87.6356	1.6	97.1283	1.3
Car	86.7359	3.6	91.9551	3.5	89.1836	3.3	94.5367	0.4
Cow	79.1248	4.4	92.8547	1.8	94.2354	3.4	94.1496	1.2
Cup	91.8082	3.0	79.9376	3.1	81.4891	2.8	91.0254	1.1
Dog	92.8834	0.1	96.7538	0.8	95.8463	2.1	93.1253	0.2
Horse	93.7593	3.9	93.7652	3.3	92.2065	1.9	95.9371	0.2
Pear	92.9417	4.2	86.6273	0.1	88.9361	2.8	96.5991	0.8
Tomato	85.9304	3.1	88.6976	1.3	89.5497	3.1	94.4962	1.7
Overall	88.9248	2.3	89.9796	2.4	89.8853	2.0	94.6247	0.9

image understanding techniques. In order to deal with the aforementioned problem and with a view to amplify the performance of our method, we expanded our training set

with images of partially occluded objects. In more detail, we artificially introduced partial occlusions in the existing training set with the percentage of obstruction lying in the range



(a) Comparative evaluation for a permissive error of 5 degrees.



(b) Comparative evaluation for a permissive error of 3 degrees.

Figure 8. By adopting the evaluation criterion of [10], we comparatively study the performance of the proposed method with respect to the percentage of artificially generated partial occlusions over the surface of the testing objects. In (a) and (b), the aggregated results in cases where a measurement of the 3D pose is considered successful if the error is less than 5° and 3° , respectively, are presented.

[0–95]. Moreover, we adopted the evaluation criterion presented in [10] so as to comparatively appraise the performance of the proposed 3D pose module against partial occlusions. The adopted evaluation metric considers a measurement of the 3D pose of an object to be acceptable when the error of the computed rotation parameters is less than 5° . Figure 8(a) depicts the superiority of our work against other related projects, whilst providing evidence of being more tolerant to partial occlusions. However, due to the fact that all methods (including ours) make use of oversampled training sets (e.g. objects shot every 5°), we have further tested the respective efficacy potentials for a permissible error of 3° , with the results being presented in figure 8(b). Once again, the proposed 3D pose manifold modeling technique incorporating an RBF kernel proved to hold occlusion invariance capacities compared to the works of Yuan *et al* [21], Hinterstoisser *et al* [10], Mei *et al* [15] and Lowe [24].

5.3. Generalization capacities

The goal of a machine learning application and, especially, of its training procedure, is to build a function that models adequately the training data. Additionally, the proposed

architecture should emphasize designing schemes that do not construct statistical models of the process that generates the data. This is how a network can maximize its generalization capacities, that is, to make adequate predictions regarding new input vectors. Over(under)fitting situations are closely related to generalization issues due to the fact that they directly affect its potential, whilst they are primarily avoided by the introduction of regularization parameters. Toward this end, in this paper, the proposed frameworks adopt regularization frameworks that minimize the possibility of over(under)fitting. Regarding the object recognition module, its generalization capacity is limited to the classes of the ETH-80 dataset, whilst as table 5.1 demonstrates, our method is capable of adequately categorizing testing samples into the corresponding classes. The proposed framework is capable of recognizing unknown objects by making use of the MI-based algorithm, which projects the samples onto a sub-space with linear separation of the different classes. Additionally, we acquire 6 DoF accurate measurements of unknown objects through the establishment of low-dimensional pose manifolds that are projected onto highly distinguishable sub-spaces. The resulting manifolds are based on a sophisticated feature extraction and representation method that, in contrast to other related projects, does not

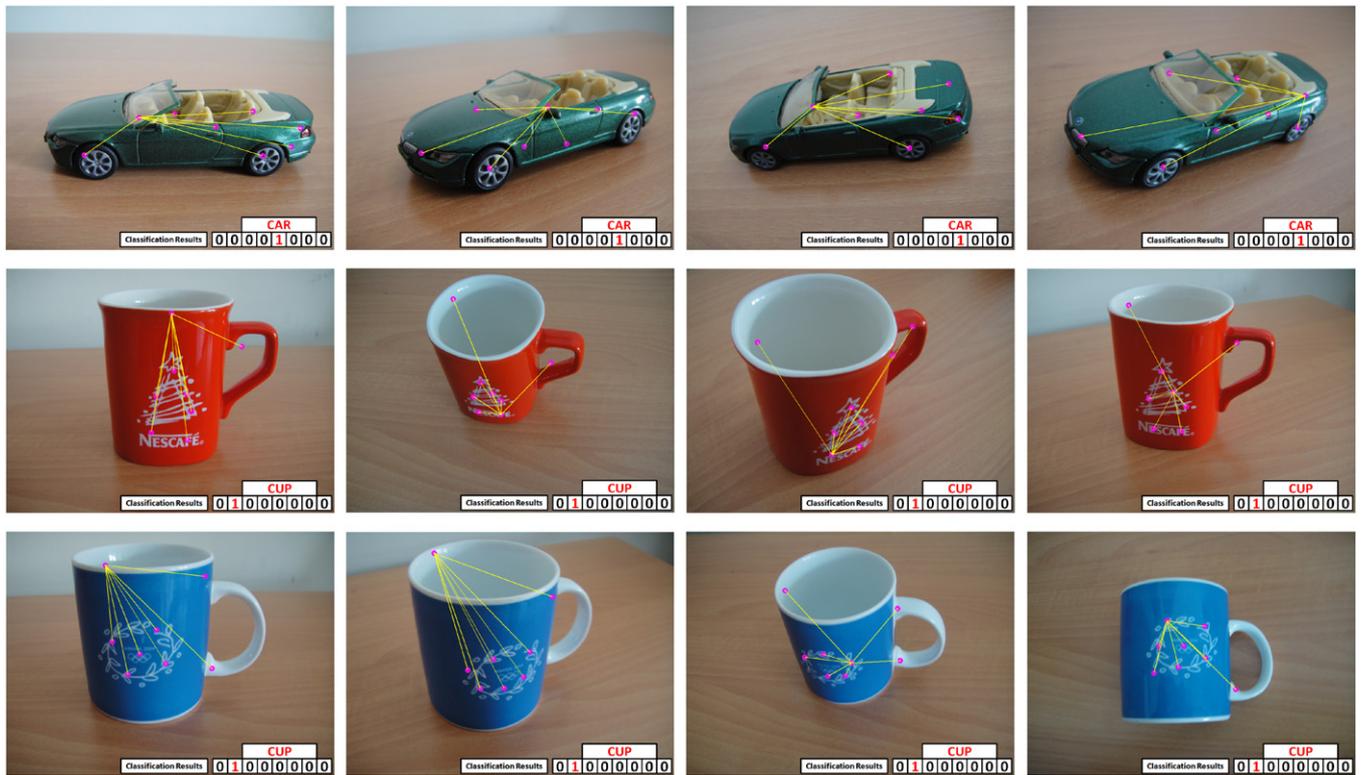


Figure 9. The proposed recognition and 3D pose estimation method is capable of generalizing to unknown objects, such as a car and two cups.

Table 2. Comparative evaluation of the generalization capacities of the algorithms. This table demonstrates the average pose estimation success for 30 unknown objects belonging to the following classes. The work of Lowe [24] is not included in the table due to the fact that it does not allow any generalization capacities.

Class	Hinterstoisser <i>et al</i> [10]		Yuan <i>et al</i> [22]		Mei <i>et al</i> [15]		Proposed method	
	Success (%)	STD	Success (%)	STD	Success (%)	STD	Success (%)	STD
Apple	89.7248	3.2	81.4248	4.2	90.6795	1.3	91.9162	1.6
Car	92.9102	2.6	90.4102	3.5	95.9082	2.1	98.3649	0.9
Cow	91.0479	2.1	87.4479	2.9	90.7792	2.2	95.0159	1.8
Cup	93.2912	1.9	87.3912	1.8	92.0512	2.6	97.2148	0.8
Dog	88.9489	3.6	79.1489	3.2	88.7489	4.3	94.6656	2.1
Horse	91.5854	2.2	84.8854	4.2	91.8114	4.1	94.8161	1.7
Pear	86.8517	4.2	81.3817	2.9	90.5217	3.4	92.7584	1.2
Tomato	84.3607	3.1	75.1607	6.2	85.9607	2.7	90.1974	0.9
Overall	89.8401	1.5	83.4064	2.5	90.8076	1.4	94.3687	1.3

rely upon the attributes of the object class and requires only minimum supervision during training. As presented in table 2, the proposed 3D pose estimation module demonstrated higher success rates, compared to related frameworks, for a permissible error of 5° . The testing subset consists of images of objects that are totally different to those of the databases used. They differ in the surrounding illumination circumstances, image resolution and the captured object-target itself (still belonging to one of the 8 classes). In figure 9, we demonstrate the generalization capability of the proposed method. We believe that our methods minimize the generalization error mainly due to the underlying manifold modeling architecture, the kernels and regularization blocks utilized.

6. Conclusion

We proposed a novel solution to the challenging tasks of simultaneously recognizing objects and providing an accurate measurement of their geometrical configuration in 3D space. Our method relies on the establishment of highly discriminative manifolds that (a) model several variations of the same object in the same loci (recognition manifolds); (b) encapsulate the basic intuition that different objects viewed under identical perspectives hold high dimensional data about their pose (3D pose manifolds). Comparative evaluation of our method against other related works of the field (a) validated our theoretical claims, (b) provided evidence of low generalization error and (c) regarding the 3D pose module, justified our choice to adopt a neural network-based strategy. According to the

methods presented by Yuan *et al* [22], Hinterstoisser *et al* [10] and Mei *et al* [15], there is evidence that the relationship between input and output spaces is a nonlinear one, while artificial neural networks are the most efficient mechanisms in modeling this liaison. Looking ahead to future work, we plan to adjust the presented modules in order to employ them in object manipulation, obstacle avoidance and other related human–robot interaction tasks.

References

- [1] Verhagen C 1982 Measurement for and by pattern recognition *J. Phys. E: Sci. Instrum.* **15** 607
- [2] Kembhavi A, Siddiquie B, Mieziako R, McCloskey S and Davis L S 2009 Incremental multiple kernel learning for object recognition *ICCV 2009: IEEE Int. Conf. on Computer Vision* pp 638–45
- [3] Bo L, Lai K, Ren X and Fox D 2011 Object recognition with hierarchical kernel descriptors *IEEE Int. Conf. on Computer Vision and Pattern Recognition* pp 1729–36
- [4] Sandhu R, Dambreville S, Yezzi A and Tannenbaum A 2009 Non-rigid 2D–3D pose estimation and 2D image segmentation *CVPR 2009: IEEE Conf. on Computer Vision and Pattern Recognition* pp 786–93
- [5] Thomas A, Ferrari V, Leibe B, Tuytelaars T and Gool L 2009 Shape-from-recognition: recognition enables meta-data transfer *Comput. Vis. Image Underst.* **113** 1222–34
- [6] Kouskouridas R, Amanatiadis A and Gasteratos A 2011 Guiding a robotic gripper by visual feedback for object manipulation tasks *IEEE Int. Conf. on Mechatronics* pp 433–8
- [7] Detry R and Piater J 2011 Continuous surface-point distributions for 3D object pose estimation and recognition *Asian Conf. on Computer Vision* pp 572–85
- [8] Kouskouridas R, Gasteratos A and Badekas E 2012 Evaluation of two-part algorithms for objects' depth estimation *IET Comput. Vision* **6** 70–8
- [9] Savarese S and Fei-Fei L 2007 3D generic object categorization, localization and pose estimation *IEEE Int. Conf. on Computer Vision* pp 1–8
- [10] Hinterstoisser S, Benhimane S and Navab N 2007 N3m: Natural 3D markers for real-time object detection and pose estimation *IEEE Int. Conf. on Computer Vision* pp 1–7
- [11] Zitov B and Flusser J 2003 Image registration methods: a survey *Image Vis. Comput.* **21** 977–1000
- [12] Chen H and Varshney P 2003 Mutual information-based CT-MR brain image registration using generalized partial volume joint histogram estimation *IEEE Trans. Med. Imaging* **22** 1111–9
- [13] Schölkopf B and Smola A J 2002 *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond* (Cambridge, MA: MIT Press)
- [14] Mei L, Sun M, Carter K M, Hero A O III and Savarese S 2009 Object pose classification from short video sequences *British Machine Vision Conf.* pp 89.1–12
- [15] Mei L, Liu J, Hero A and Savarese S 2011 Robust object pose estimation via statistical manifold modeling *Int. Conf. on Computer Vision* pp 967–74
- [16] Kyriakoulis N and Gasteratos A 2010 Color-based monocular visuo-inertial 3D pose estimation of a volant robot *IEEE Trans. Instrum. Meas.* **59** 2706–15
- [17] Payet N and Todorovic S 2011 From contours to 3D object detection and pose estimation *ICCV 2011: IEEE Int. Conf. on Computer Vision* pp 983–90
- [18] Elqursh A and Elgammal A 2011 Line-based relative pose estimation *CVPR 2011: IEEE Conf. on Computer Vision and Pattern Recognition* pp 3049–56
- [19] Chinellato E and del Pobil A 2009 Distance and orientation estimation of graspable objects in natural and artificial systems *Neurocomputing* **72** 879–86
- [20] Nian R, Ji G, Zhao W and Feng C 2007 Probabilistic 3D object recognition from 2D invariant view sequence based on similarity *Neurocomputing* **70** 785–93
- [21] Yuan C and Niemann H 2001 Neural networks for the recognition and pose estimation of 3D objects from a single 2D perspective view *Image Vis. Comput.* **19** 585–92
- [22] Yuan C and Niemann H 2003 Neural networks for appearance-based 3D object recognition *Neurocomputing* **51** 249–64
- [23] Fergus R, Perona P and Zisserman A 2007 Weakly supervised scale-invariant learning of models for visual recognition *Int. J. Comput. Vis.* **71** 273–303
- [24] Lowe D 1999 Object recognition from local scale-invariant features *Int. Conf. on Computer Vision* **2** 1150–7
- [25] Bailly K and Milgram M 2009 Boosting feature selection for neural network based regression *Neural Netw.* **22** 748–56
- [26] Jolliffe I 1986 *Principal Component Analysis* (Berlin: Springer)
- [27] Duda R O, Hart P E and Stork D G 2001 *Pattern Classification* 2nd edn (New York: Wiley)
- [28] Hyvärinen A and Oja E 2000 Independent component analysis: algorithms and applications *Neural Netw.* **13** 411–30
- [29] Schölkopf B, Smola A J and Müller K R 1997 Kernel principal component analysis *Int. Conf. on Artificial Neural Networks* pp 583–8
- [30] Baudat G and Anouar F 2000 Generalized discriminant analysis using a kernel approach *Neural Comput.* **12** 2385–404
- [31] Belkin M and Niyogi P 2003 Laplacian eigenmaps for dimensionality reduction and data representation *Neural Comput.* **15** 1373–96
- [32] Roweis S T and Saul L K 2000 Nonlinear dimensionality reduction by locally linear embedding *Science* **290** 2323–6
- [33] Balasubramanian M, Shwartz E L, Tenenbaum J B, de Silva V and Langford J C 2002 The isomap algorithm and topological stability *Science* **295** 7a
- [34] Yan S, Xu D, Zhang B, Zhang H J, Yang Q and Lin S 2007 Graph embedding and extensions: a general framework for dimensionality reduction *IEEE Trans. Pattern Anal. Mach. Intell.* **29** 40–51
- [35] He X and Niyogi P 2003 Locality preserving projections *NIPS*
- [36] Walters-Williams J and Li Y 2009 Estimation of mutual information: a survey *Proc. 4th Int. Conf. on Rough Sets and Knowledge Technology* (Berlin: Springer) pp 389–96
- [37] Lu X, Zhang S, Su H and Chen Y 2008 Mutual information-based multimodal image registration using a novel joint histogram estimation *Comput. Med. Imaging Graph.* **32** 202–9
- [38] Plum J P W, Maintz J B A and Viergever M A 2003 Mutual-information-based registration of medical images: a survey *IEEE Trans. Med. Imaging* **986**–1004
- [39] Meila M 2003 Comparing clusterings by the variation of information *Lecture Notes in Computer Science* (Berlin: Springer) pp 173–87
- [40] Lowe D 2004 Distinctive image features from scale-invariant keypoints *Int. J. Comput. Vis.* **60** 91–110
- [41] Fischler M and Bolles R 1981 Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography *Commun. ACM* **24** 381–95
- [42] Bishop C 2006 *Pattern Recognition and Machine Learning* (New York: Springer)
- [43] Rodriguez J D, Perez A and Lozano J A 2010 Sensitivity analysis of k-fold cross validation in prediction error estimation *IEEE Trans. Pattern Anal. Mach. Intell.* **32** 569–75

- [44] Nayar S, Nene S and Murase H 1996 Columbia object image library (coil 100) *Technical report no. CUCS-006-96* Department of Computer Science, Columbia University
- [45] Evermotion T M Evermotion 3D models <http://www.evermotion.org/>
- [46] Viksten F, Forssen P, Johansson B and Moe A 2009 Comparison of local image descriptors for full 6 degree-of-freedom pose estimation *ICRA'09: IEEE Int. Conf. on Robotics and Automation* pp 2779–86
- [47] Pang Y, Li W, Yuan Y and Pan J 2012 Fully affine invariant surf for image matching *Neurocomputing* **85** 6–10
- [48] Arulselvi M, Ramalingam V and Palanivel S 2011 Article: cephalometric analysis using PCA and SVM *Int. J. Comput. Appl.* **30** 39–47
- [49] Avci E 2012 A new expert system for diagnosis of lung cancer: GDA LS SVM *J. Med. Syst.* **36** 2005–9
- [50] Yang J 2011 Language recognition with locality preserving projection *ICDT 2011: 6th Int. Conf. on Digital Telecommunications* pp 51–55
- [51] Duan K and Keerthi S S 2005 Which is the best multiclass SVM method? an empirical study *Proc. 6th Int. Workshop on Multiple Classifier Systems* pp 278–85
- [52] Vapnik V N 1998 *Statistical Learning Theory* (New York: Wiley-Interscience)