Pose Manifolds for Efficient Visual Servoing

Rigas Kouskouridas, Angelos Amanatiadis and Antonios Gasteratos
Democritus University of Thrace
Department of Production and Management Engineering
Vas. Sofias 12, Building I, Xanthi Greece 67100
Emails: rkouskou@pme.duth.gr aamanat@ee.duth.gr gasteratos@ieee.org

***Abstract*—In order to adequately accomplish vision-based manipulation tasks, robotic platforms require an accurate estimation of the 3D pose of the target, which is efficiently approached by imaging techniques excessively utilizing large databases that consist of images of several objects captured under varying viewpoints. However, such approaches are characterized by large computational burden and complexity accompanied by limited capacities to interpolate between two known instances of an object. To address these issues we propose a robust 3D object pose estimation technique that entails a manifold modeling procedure based on appearance, geometrical and shape attributes of objects. We utilize a bunch-based method that is followed by a shape descriptor module, in order to establish low dimensional pose manifolds capable of distinguishing similar poses of different objects into the corresponding classes. Finally, an accurate estimation of the 3D pose of a target is provided by a neural network-based solution that encompasses a novel input-output space targeting method. We have comparatively studied the performance of our method against other related works, whilst experimental results justify our theoretical claims and provide evidence of low generalization error.**

I. Introduction

Recent research endeavors in the field of robotics and mechatronics were dedicated to the designing and implementation of advanced robotic platforms that aim at minimizing user's effort. Towards this end, new generation frameworks are primarily equipped with advanced visual sensors that enable the autonomous grasping of objects in the working envelope of the robot. Object manipulation stands for the most widely realized operation of a human being since it is directly linked to several vital tasks, i.e. eating or drinking. According to the literature, manipulation frameworks could be categorized into four major sub-classes depending on simplicity, reliability and versatility levels of the corresponding frameworks. The first category represents mechatronics schemes, referred as workstations, which depend upon position-based grasping modules that are mainly comprised of a robotic arm fixed to a desk [1], [2], [3], [4]. The main attribute characterizing the systems belonging to the second category is the introduction of sensors' feedback. Additionally, stand alone manipulators are usually lighter than workstations, whilst the total knowledge of their working environment does not constitute a prerequisite [5], [6]. Wheel chair mounted frameworks that correspond to systems associated with the third sub-category, emphasize in increasing the working space of the robotic arm by mounting the latter on a moving particle [7], [8], [9]. In the last category the most advanced assistive robots [10], [11] are assorted, which are primarily comprising of a robotic arm mounted onto an autonomously moving vehicle equipped with either laser [12], [13] or visual [14], [15] sensors.

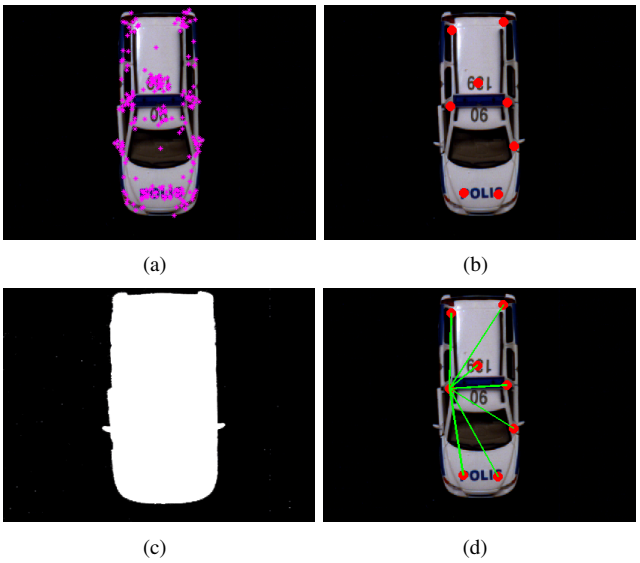

Fig. 1. In this figure the main idea underlying the proposed method is shown. The part-based architecture is built through the processes of key-point extraction 1(a) and a clustering procedure 1(b). Each member of the proposed bunch-based structure encapsulates both appearance and geometrical attributes of the object. The next phase encompasses the process of the shape extraction 1(c). The establishment of the manifold is adequately accomplished through the calculation of the distance of each of these clusters with a given one 1(d). As a final step, several manifolds representing numerous training objects are considered as input to a neural network-based framework that is simulated in order to estimate the 3D pose of the object to be given as an input to the manipulator.

Visual servoing, which has been widely used to solve for the object manipulation problem, incorporates among others a kind of 3D object pose estimation task. The latter stands for one of the most challenging problems in computer vision, mainly due to its practical caliber and its ability to be adopted into a plethora of various applications. Generally, a 3D pose estimation algorithm embodies sophisticated routines that aim at pledging vital visual information regarding the geometrical configuration of an object-target [16], [17], [18], [19], [20], [21]. In its most general form, the 3D pose estimation problem is solved by the efficient exploitation of trained databases that imply efficient solution to the 2D-2D image feature correspondence problem. However, such approaches

978-1-4577-1775-8/12/$26.00 ©2012 IEEE

are characterized by high complexity accompanied by large computational burden.

In this paper we propose a visual servoing approach, which is based on a manifold modeling module, and that it could be easily adopted by any advanced mechatronics platform. The proposed manifold building process is based on the intuition that similar poses of different objects captured under identical viewpoints share akin manifolds. The latter are established through an advanced process that comprises of a part-based structure enclosing both appearance and geometrical attributes of the trained objects. One could say that the manifold of the 3D model of an object in a known training instance is governed by the distances of the members of the bunch-based architecture from one particular center as shown in Fig. I. An accurate estimation of the 3D pose of a testing object is obtained through the simulation of a neural network that is trained with several available datasets. The obtained 3D object pose estimation is given as an input to the gripper in order to adequately fulfill object manipulation tasks. The contribution of this paper, among others, entails the formalization of this novel manifold modeling that avoids the use of conventional dimensionality reduction techniques widely used in computer vision and robotics applications. Additionally, the proposed part-based implementation is capable of pledging both appearance and geometrical attributes of the objects. Moreover, the extracted manifolds are of low dimensionality, which in turn, avoids the use of conventional dimensionality reduction schemes. Furthermore, the proposed neural network-based solution encompasses a new input-output space mapping that directly establishes a liaison between input and target spaces that does not depend upon dimensionality adjustment modules. The performance of the proposed method has been comparatively studied against other related projects for 3D object pose estimation that are based on a) manifold modeling, b) part-based solutions, c) conventional Principal Component Analysis (PCA) and d) least-squares solution. Experimental results provide evidence of low generalization error whilst justifying our theoretical claims and our choice to adopt a neural network-based solution. Finally, the most important attribute of the proposed approach constitutes the fact that it can be easily adopted by any advanced robotic framework that either aims at solving the 3D pose estimation problem or is dedicated to object manipulation processes.

II. Related work

Our work takes advantage of previous research conducted in the area of feature representation and extraction for 3D object recognition and pose estimation [22], [23], [24], [25]. Picking up the most prominent patterns represents a challenging task with a significant effect on the 2D-3D point correspondence. In [26], a new method for efficient feature selection based on a Fuzzy Functional Criterion, for the evaluation of the linkage between the input features and the output score, is presented. However, this technique is dedicated, specifically, to the head pose problem for which it can report remarkable efficiency when trained with numerous datasets. On the other hand, in [27] and its extension [28] it was shown that a compact model of an object can be portrayed by linking together diagnostic parts of the objects from different viewpoints. Such parts are defined as large and distinguishable regions of the objects that are composed of many local invariant features. Despite the efficiency of this architecture, the method is mainly devoted to 3D object categorization.

Although neural networks are common place in several computer vision applications, for the particular task of estimating the 3D pose of an object only few approaches have been proposed. Early studies [29], [30], [31], [32], [33] showed that the adoption of neural networks in image processing is recommended in cases where the task in hand encompasses great physical complexity. In more detail, a modification of Kohonens self-organizing feature map (SOM) is trained with computer generated object views corresponding to one or more object orientation parameters. Although the methods presented in those papers reported significant gains in performance, their networks could achieve generalizations for trained objects only. Several methods in this area adopt dimensionality reduction schemes with the PCA and its variations being the most popular one. For instance, in [34] an appearance-based method for the efficient estimation of the pose of 3D objects, where the PCA is utilized for dimensionality reduction, is presented. The neural network is trained with the resilient backpropagation method and, as far as the rotation parameters of the pose are concerned, only two DoFs are estimated, corresponding to in and out of image plane orbits. An extension of the aforementioned technique is found in [35] where input feature vectors are derived by nonlinear PCA. Both methods fail to interpolate between two known pose configurations, since they utilize object views with a sampling interval of 3^o and emphasize in distinguishing the input patterns into the corresponding classes.

In [36] and [28] the closest works to our paper, regarding the feature extraction and manifold modeling processes, are presented. In particular in [36] a method for finding 4 or 5 close feature points, called Natural 3D Markers (N3Ms), is presented, whilst the extracted features enjoy distinctive photometric properties and equal distribution over the objects visible surface. While the two approaches, both ours' and the one presented in [36], have similarities we believe our model provides a more compact and abstract representation of the 3D object. On the other hand, in [28] a method for efficient manifold modeling that enables accurate 3D object pose estimation is presented. Despite its sufficient results, this approach has built manifolds of high dimensionality that do not credit for intra-class distance minimization and inter-pose variability maximization, respectively. In our case, we address these issues by employing a modeling process that establishes manifolds of low dimensionality that are augmented by a shape descriptor. Moreover, we show that the proposed input-output mapping adopted by a neural network strategy is experimentally proved to provide more accurate results.

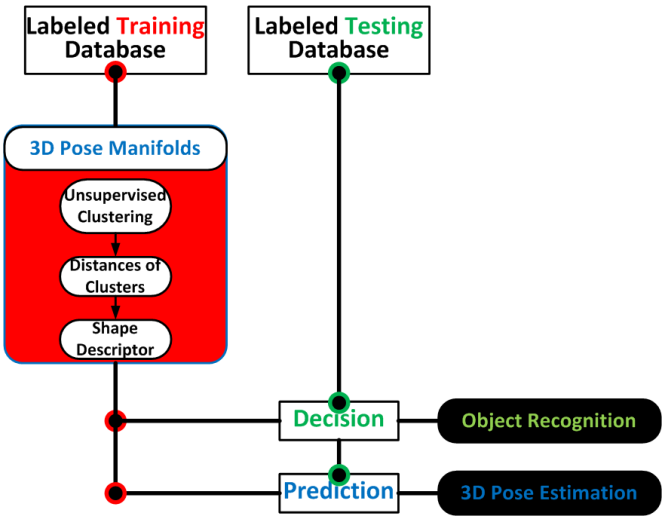

Fig. 2. The first stage of our approach incorporates the division of the labeled databases into the the testing and training subsets. During training images of objects with altering pose are given as input to the sub-routine of manifold modeling that entails a) the unsupervised clustering of the abstracted features, b) the subroutine of manifold establishment and c) the shape descriptor utilized for both 3D pose estimation and recognition purposes. The ultimate goal of our method is provide accurate decisions and predictions regarding the class of the object and its pose, respectively.

III. Methodology

The proposed visual servoing method consists of into two major components that directly affect the overall performance of the system. The first module is dedicated to object recognition that makes use of a unified database based on MPEG-7 [37] and ETH-80 [38] binary shape datasets. MPEG-7 dataset was utilized during the Core Experiment CE-Shape-1, part B and contains 70 object classes with 20 images per class resulting in 1400 images of objects. Additionally, the aforementioned database is adopted due to the fact that it is very useful for testing the performance of similarity-based retrieval and the efficacy of shape descriptors. Moreover, ETH-80 dataset represents an integrated database of images of 8 different categories of objects and comprises of 80 targets captured under 41 different viewpoints. During the testing session of our framework, where an unknown object is provided to the system, the shape classification algorithm categorizes the unknown particle to the corresponding classes.

The second module focuses on providing accurate measurements regarding the 3D pose of the testing object. Highly motivated from the fact that objects with identical poses lie on similar subspaces, we propose a manifold modeling procedure that makes use of two databases [39], [40] that contain collections of different targets viewed under altering viewpoints. Additionally, we enhance our training dataset by introducing synthetically rendered data, where 3D models (available in [41]) are captured under different geometrical configurations of the camera. The 3D pose estimation module is based on the unsupervised clustering of the extracted keypoints that unlike in [28] does not depend on the "alignment" and "expansion" operations. Furthermore, our manifold modeling framework imposes upon a bunch-based architecture that encapsulates both geometrical and appearance-based characteristics of the objects. Additionally, due to the fact that our system makes use of several databases containing numerous objects, our 3D pose estimation module provides efficient predictions regardless of the object category.

A. Bunch-based Architecture and Manifold Modeling

The ultimate scope of this subroutine is to design and implement a part-based formation capable of distinctively encapsulating several attributes of the objects, to be fed into the manifold modeling process. The proposed bunch-based structure abstracts bunches or parts that enjoy both appearance-based characteristics and geometrical ones, i.e. their location distribution. In [28] such bunches are abstracted in a supervised manner, meaning that each part stands for the realization of the Probability Density Function (PDF) associated with the joint distribution of appearance and geometry attributes of a target. Notwithstanding their sufficient performance levels, such an approach, lacks generalization capacities to unknown objects, since it can sufficiently represent only object models that are already included in the training dataset. In this paper, we deal with this problem by incorporating an unsupervised clustering that is employed over the extracted keypoints of the object. As far as the appearance-based characteristics are concerned the latter are acquired by extracting invariant features organized in high dimensional vectors by the SIFT descriptor [42] followed by homography-based RANSAC [43] for outliers removal . As a follow-up step, topological characteristics of the parts are aggregated by a K-means unsupervised clustering method. Let $I(\chi, O_i|P_j)$ representing the raw intensity values for the three-channeled image I of object $O_i \in [O_1, O_2, \ldots, O_m]$ with pose $P_j \in [P_1, P_2, \ldots, P_n]$. We then extract ρ interest key-points and denote them as $I(\boldsymbol{x}^\rho, O_i|P_j)^d$, $\boldsymbol{x} \in \mathbb{R}^2$, 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 resulted vector is normalized to unit length, in order to maintain invariance to affine changes in illumination. As a follow-up step, we assume only the locations $\boldsymbol{x} \in \mathbb{R}^2$ of the extracted appearance-based features and we find those clusters that minimize the following objective function:

$$J(c^{(1)}, \ldots, c^{(\rho)}, \mu^{(1)}, \ldots, \mu^{(\kappa)}) = \frac{1}{\rho} \sum_{i=1}^{\rho} ||\boldsymbol{x}^i - \mu_{c^{(i)}}||^2 \quad (1)$$

The construction of our bunch-based architecture is demonstrated in the pseudocode form 1.

B. Manifold Modeling

In this paper as a manifold $\mathcal{M}$ we define a locally Euclidean topological space that is governed by a continuous and invertible function $\phi(u)$ mapping any point u $\in \mathcal{M}$ to a point $v = [V_1 ... V_d]^T \in \mathbb{R}^d$. d corresponds to the dimensionality of the manifold $\mathcal{M}$ and $v = [V_1 ... V_d]^T$ to the established coordinate system. It is palpable that the most important attribute of a manifold is the one of quantizing input data that in most of

Algorithm 1 Calculate the positions μ_κ of the γ clusters

Inputs: Locations of the ρ extracted SIFT features of the object O_i with pose P_j
while $||J(t+1) - J(t)|| > \epsilon$ **do**
    Compute:
    $J(c^{(1)}, \ldots, c^{(\rho)}, \mu^{(1)}, \ldots, \mu^{(\kappa)}, t) \leftarrow \frac{1}{\rho} \sum_{i=1}^{\rho} ||\boldsymbol{x}^i - \mu_{c^{(i)}}||^2$
    $t = t + 1$
end while
where:
    $c^{(\rho)}$ = index of cluster (1,2, ...,κ) to which feature ρ at location $\boldsymbol{x}^\rho$ is initially randomly assigned
    μ_κ = position of cluster centroid κ ($\mu_\kappa \in \mathbb{R}^2$)
    $\mu_{c^{(\rho)}}$ = centroid of cluster to which example $\boldsymbol{x}^\rho$ is assigned after one step of the algorithm
Outputs: Locations of the γ clusters over the surface of the object

the times are of high dimensionality. However, in the particular problem of 3D object pose estimation, particular emphasis should be given in the dimensionality of the manifold $\mathcal{M}$, since the latter affects directly the efficiency of any regressor or classifier trained with large datasets. Towards this end, the ultimate goal of the proposed work was to find the function $\phi(u)$ that establishes a manifold $\mathcal{M}'$ with the minimum potential dimensionality. Additionally, the resulting manifold $\mathcal{M}'$ should be capable of efficiently categorizing similar poses of several objects captured under identical viewpoints into the same pose space. Towards this end, the establishment of the manifold is adequately fulfilled via the calculation of the distance of each member of the bunch-based architecture to a given one.

Regarding the specific training object $(O_i^*|P_j^*)$ we assume that the computed members of the bunch-based architecture 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 framework proceeds by computing the L_2 norm between vector $\mathbf{e}_i \in \mathcal{E}$ and anchor point $\mathbf{e}^*$:

$$\mathbf{x}^i = ||\mathbf{e}_i - \mathbf{e}^*||^2 = \sum_{i=1}^{c} \{\mathbf{e}_i - \mathbf{e}^*\}^2 \tag{2}$$

Afterwards, we train a Radial Basis Functions-based regressor in order to acquire a mapping from a set of input variables $\mathbf{x} = [x_1, \ldots, x_d]$, belonging to a feature-space $\mathcal{X}$, to a modeled output variable $\mathbf{y} = \mathbf{y}(\mathbf{x}; w) \in \mathcal{Y}$, with w denoting the vector of the adjustable parameters. The ultimate goal of our system is to learn a regressor $g : \mathcal{X} \rightarrow \mathcal{Y}$ from an a priori training dataset $\{\mathbf{x}^n, \mathbf{y}^n\}$, in order to efficiently approximate the output $\mathcal{Y}_t$, when an unknown example $\mathcal{X}_t$ is provided. The proposed architecture encompasses a new input-output mapping procedure that does not require common dimensionality reduction operations. In particular, opposed to [28] and [34] where the input vectors are of very high dimensionality, the established manifolds are of low one and are directly fed into the regressor.

C. Shape Retrieval and Classification

The ultimate goal of this module is to extract accurate and highly representative shape-related information. In cases where invariance with respect to a number of possible transformations, such as scaling, shifting, and rotation, is required, the task of extracting the respective shape descriptors is characterized by large complexity and computational burden [44]. In this paper we address this issue by utilizing a complementary scheme of three different shape descriptors for achieving optimum accuracy in shape representation [45]. Particularly, we employ the Fourier descriptors [46], which are contour-based schemes that enjoy contour properties of the object, and both affine moment invariants [47] and region-based angular radial transform descriptors [48]. We revised the state-of-the-art literature in the field of shape descriptors in order to set the optimal number of coefficients with a view to provide the pillars of efficient descriptor indexing. In more detail, regarding the Fourier descriptors we utilize only the first 32 descriptors, for the angular radial transform the first 35 and 6 affine moment invariants. Object recognition is achieved through the Fuzzy Lattice Reasoning (FLR) scheme as outlined in a preliminary work [49]. In [49], a 2-D shape is represented in a form of three populations of three different shape descriptors $d \in FD, ART, IM$, respectively, such that one population to corresponds to one Intervals' Number (IN) as detailed in [50], [51]. Additionally, a FLRtypeII scheme for learning (training) was applied followed by a FLRtypeII scheme for generalization (testing) on the benchmark data set. Unknown object categorization is adequately accomplished by classifying the testing target into the class with the smallest distance.

IV. Experimental Results

The performance of the proposed manifold modeling approach was evaluated through a series of experiments that a) are dedicated to the 3D object pose estimation problem and b) focus on object manipulation. Regarding the first category of experiments particular emphasis is given to partial occlusions since the latter affect directly the efficiency of any computer vision scheme. Although several research endeavors and achievements reached so far, an advanced

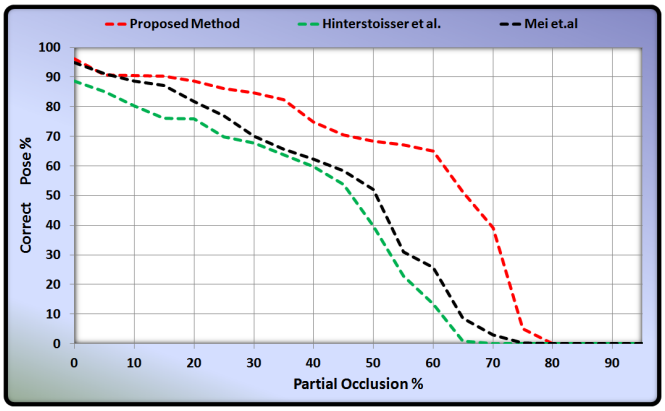

Fig. 3. The performance of our method against partial occlusions is comparatively evaluated with the works of Hinterstoisser et al. [36] and Mei et al. [28].

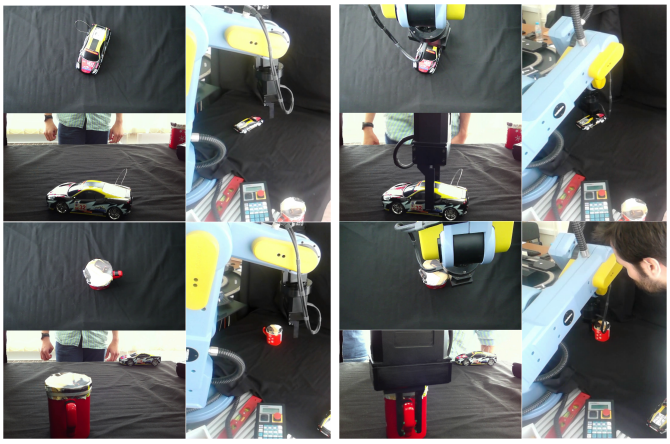

Fig. 4. During these series of experiments we utilized three vision sensors, one installed over the working space of the robot, one capturing the working space sideways and one overseeing the overall procedure. Additionally, we utilize the SCORBOTE-ER Vplus robotic arm, which is a vertical articulated robot with 6 DoFs, in order perform object manipulation tasks.

imaging method typified with adequate trade-offs between computational burden and high generalization capacities, has yet to been built. In our work the problem induced by partial occlusions is addressed by expanding our training database with images of partial occluded objects. Concretely, partially occluded objects that are introduced in the existing database with the percentage of artificially generated obstruction lying in the range [0-95]. Additionally, we adopt the evaluation criterion presented in [36] in order to comparatively evaluate the performance of the proposed 3D pose estimation module against partial occlusions. According to the particular metric of [36] a measurement regarding the 3D pose of an object is considered as successful in cases where the error of the computed rotation parameters is less than 5^0. The superiority of our work compared to other related works is illustrated in Fig. 3. It is palpable that, experimental results of this first category provide evidence that our method is more tolerant to partial occlusions than the works of Hinterstoisser et.al [36] and Mei et al.[28].

We have additionally evaluated the performance of the proposed framework through experiments dedicated to object manipulation. According to the literature, in the field of robotics research two major camera configurations for vision-based scene understanding are discerned. Eye-in-hand and eye-to-hand architectures entail either mounting the vision sensor(s) onto the robot's end-effector or installing cameras capable of observing the working space of the robot, respectively. In this paper we propose an eye-to-hand camera configuration without, however, limiting our work to such scenarios. Indeed, initial experimental results demonstrated that our manifold modeling approach that utilizes shape information of objects, can be easily adopted to accomplish object manipulation tasks within an eye-in-hand architecture. Fig. 4 demonstrates the efficient accomplishment of several manipulation tasks by the proposed method.

V. Conclusion

In this paper we presented a novel solution to the problem of automatic vision-based object manipulation. Our framework lays its foundations on the intuition that different objects viewed under the same perspective share identical poses that can be efficiently projected onto high representative subspaces. We employ a part-based scheme that encapsulates several attributes of the objects such as shape, appearance and geometry. Our manifold modeling procedure imposes upon the unsupervised clustering of the extracted visual cues that is responsible for feeding an RBF-based regressor. Comparative experimental results demonstrated a) the invariance of our work against partial occlusions and b) the superiority of our work opposed to other related projects. Regarding the future work, we aim at designing and implementing a novel manifold modeling architecture capable of creating without supervision ontology-based databases that hold information exploited in challenging image understanding tasks.

References

[1] M. Busnel, R. Cammoun, F. Coulon-Lauture, J. Détriché, G. Le Claire, and B. Lesigne, "The robotized workstation "MASTER" for users with tetraplegia: Description and evaluation," *Journal of rehabilitation research and development*, vol. 36, no. 3, pp. 217–229, 1999.

[2] R. Soyama, S. Ishii, and A. Fukase, "The development of mealassistance robotmyspoon," in *Proceedings of the 8th International Conference on Rehabilitation Robotics*, 2003, pp. 88–91.

[3] M. Topping, "An overview of the development of Handy 1, a rehabilitation robot to assist the severely disabled," *Journal of Intelligent & Robotic Systems*, vol. 34, no. 3, pp. 253–263, 2002.

[4] H. F. M. Van der Loos, J. J. Wagner, N. Smaby, K. Chang, O. Madrigal, L. J. Leifer, and O. Khatib, "ProVAR assistive robot system architecture," in *IEEE International Conference on Robotics and Automation*, vol. 1. IEEE, 1999, pp. 741–746.

[5] A. Casals, R. Villa, and D. Casals, "A soft assistant arm for tetraplegics," in *Rehabilitation technology: strategies for the European Union: proceedings of the 1st TIDE Congress*, 1993, p. 103.

[6] K. Kawamura, S. Bagchi, M. Iskarous, and M. Bishay, "Intelligent robotic systems in service of the disabled," *Rehabilitation Engineering, IEEE Transactions on*, vol. 3, no. 1, pp. 14–21, 2002.

[7] R. M. Alqasemi, E. J. McCaffrey, K. D. Edwards, and R. V. Dubey, "Wheelchair-mounted robotic arms: Analysis, evaluation and development," in *Advanced Intelligent Mechatronics. Proceedings, 2005 IEEE/ASME International Conference on*. IEEE, 2005, pp. 1164–1169.

[8] I. Volosyak, O. Ivlev, and A. Graser, "Rehabilitation robot FRIEND II-the general concept and current implementation," in *Rehabilitation Robotics, 2005. ICORR 2005. 9th International Conference on*. IEEE, 2005, pp. 540–544.

[9] F. Bley, M. Rous, U. Canzler, and K. F. Kraiss, "Supervised navigation and manipulation for impaired wheelchair users," in *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, vol. 3. IEEE, 2005, pp. 2790–2796.

[10] A. Remazeilles, C. Leroux, and G. Chalubert, "SAM: a robotic butler for handicapped people," in *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on*. IEEE, 2008, pp. 315–321.

[11] Z. Bien, D. J. Kim, M. J. Chung, D. S. Kwon, and P. H. Chang, "Development of a wheelchair-based rehabilitation robotic system (KARES II) with various human-robot interaction interfaces for the disabled," in *Advanced Intelligent Mechatronics, 2003. AIM 2003. Proceedings. 2003 IEEE/ASME International Conference on*, vol. 2. IEEE, 2003, pp. 902–907.

[12] G. Bolmsjo and H. Neveryd, "WALKY, an Ultrasonic Navigating Mobile Robot for the Disabled," in *Proceedings of the 2nd TIDE Congress*, 1995, pp. 366–370.

[13] B. Graf, M. Hans, and R. D. Schraft, "Care-O-bot IIDevelopment of a next generation robotic home assistant," *Autonomous robots*, vol. 16, no. 2, pp. 193–205, 2004.

[14] R. Bischoff and V. Graefe, "Machine vision for intelligent robots," in *IAPR Workshop on Machine Vision Applications. Makuhari/Tokyo*, 1998, pp. 167–176.

[15] P. Hoppenot and E. Colle, "Localization and control of a rehabilitation mobile robot by close human-machine cooperation," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 9, no. 2, pp. 181–190, 2002.

[16] A. Thomas, V. Ferrari, B. Leibe, T. Tuytelaars, and L. V. Gool, "Shape-from-recognition: Recognition enables meta-data transfer," *Computer Vision and Image Understanding*, vol. 113, no. 12, pp. 1222 – 1234, 2009.

[17] R. Sandhu, S. Dambreville, A. Yezzi, and A. Tannenbaum, "Non-rigid 2d-3d pose estimation and 2d image segmentation," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 786 – 793, 2009.

[18] S. Ekvall, D. Kragic, and F. Hoffmann, "Object recognition and pose estimation using color cooccurrence histograms and geometric modeling," *Image and Vision Computing*, vol. 23, no. 11, pp. 943 – 955, 2005.

[19] R. Kouskouridas, E. Badekas, and A. Gasteratos, "Simultaneous visual object recognition and position estimation using SIFT," *International Conference on Intelligent Robotics and Applications*, pp. 866 – 875, 2009.

[20] R. Kouskouridas, A. Gasteratos, and E. Badekas, "Evaluation of two-part algorithms for objects' depth estimation," *Computer Vision, IET*, vol. 6, no. 1, pp. 70–78, 2012.

[21] R. Kouskouridas, A. Amanatiadis, and A. Gasteratos, "Guiding a robotic gripper by visual feedback for object manipulation tasks," in *Mechatronics (ICM), 2011 IEEE International Conference on*. IEEE, 2011, pp. 433–438.

[22] D. Lowe, "Object recognition from local scale-invariant features," *International Conference on Computer Vision*, vol. 2, pp. 1150–1157, 1999.

[23] A. Berg, T. Berg, and J. Malik, "Shape matching and object recognition using low distortion correspondences," *International Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 26–33, 2005.

[24] R. Fergus, P. Perona, and A. Zisserman, "Weakly supervised scale-invariant learning of models for visual recognition," *International Journal of Computer Vision*, vol. 71, no. 3, pp. 273–303, 2007.

[25] B. Leibe, A. Leonardis, and B. Schiele, "Combined object categorization and segmentation with an implicit shape model," *Workshop on Statistical Learning in Computer Vision European Conference on Computer Vision*, vol. 1, pp. 17–32, 2004.

[26] K. Bailly and M. Milgram, "Boosting feature selection for neural network based regression," *Neural Networks*, vol. 22, no. 5-6, pp. 748–756, 2009.

[27] S. Savarese and L. Fei-Fei, "3d generic object categorization, localization and pose estimation," *International Conference on Computer Vision*, vol. 1, pp. 1–8, 2007.

[28] L. Mei, J. Liu, A. Hero, and S. Savarese, "Robust object pose estimation via statistical manifold modeling," *International Conference on Computer Vision*, 2011.

[29] S. Winkler, "Model-based pose estimation of 3-d objects from camera images by using neural networks," *Technical Report 515-96-12, Master's Thesis*, 1996.

[30] P. Wunsch, S. Winkler, and G. Hirzinger, "Real-time pose estimation of 3d objects from camera images using neural networks," *International Conference on Robotics and Automation*, vol. 4, pp. 3232–3237, 1997.

[31] E. Chinellato and A. del Pobil, "Distance and orientation estimation of graspable objects in natural and artificial systems," *Neurocomputing*, vol. 72, no. 4-6, pp. 879–886, 2009.

[32] W. Li and T. Lee, "Projective invariant object recognition by a hopfield network," *Neurocomputing*, vol. 62, pp. 1–18, 2004.

[33] R. Nian, G. Ji, W. Zhao, and C. Feng, "Probabilistic 3d object recognition from 2d invariant view sequence based on similarity," *Neurocomputing*, vol. 70, no. 4-6, pp. 785–793, 2007.

[34] C. Yuan and H. Niemann, "Neural networks for the recognition and pose estimation of 3d objects from a single 2d perspective view," *Image and Vision Computing*, vol. 19, no. 9-10, pp. 585–592, 2001.

[35] C. Yuann and H. Niemann, "Neural networks for appearance-based 3-d object recognition," *Neurocomputing*, vol. 51, pp. 249–264, 2003.

[36] S. Hinterstoisser, S. Benhimane, and N. Navab, "N3m: Natural 3d markers for real-time object detection and pose estimation," *International Conference on Computer Vision*, vol. 1, pp. 1–7, 2007.

[37] L. Gorelick, M. Galun, E. Sharon, R. Basri, and A. Brandt, "Shape representation and classification using the poisson equation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 12, pp. 1991–2005, 2006.

[38] B. Leibe and B. Schiele, "Analyzing appearance and contour based methods for object categorization," in *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, vol. 2. IEEE, 2003, pp. II–409.

[39] F. Viksten, P. Forssen, B. Johansson, and A. Moe, "Comparison of local image descriptors for full 6 degree-of-freedom pose estimation," *ICRA*, pp. 2779–2786, 2009.

[40] S. Nayar, S. Nene, and H. Murase, "Columbia object image library (coil 100)," Tech. Report No. CUCS-006-96. Department of Comp. Science, Columbia University, Tech. Rep., 1996.

[41] T. M. Evermotion, "Evermotion 3d models," http://www.evermotion.org/.

[42] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[43] M. Fischler and R. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[44] S. Belongie, J. Malik, and J. Puzicha, "Shape Matching and Object Recognition Using Shape Contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 509–522, 2002.

[45] A. Amanatiadis, V. G. Kaburlasos, A. Gasteratos, and S. E. Papadakis, "A comparative study of invariant descriptors for shape retrieval," in *Imaging Systems and Techniques, 2009. IST'09. IEEE International Workshop on*. IEEE, 2009, pp. 391–394.

[46] D. Zhang and G. Lu, "Shape-based image retrieval using generic Fourier descriptor," *Signal Processing: Image Communication*, vol. 17, no. 10, pp. 825–848, 2002.

[47] J. Flusser and T. Suk, "Pattern recognition by affine moment invariants," *Pattern Recognition*, vol. 26, no. 1, pp. 167–174, 1993.

[48] M. Bober, "MPEG-7 visual shape descriptors," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 11, no. 6, pp. 716–719, 2001.

[49] V. G. Kaburlasos, A. Amanatiadis, and S. E. Papadakis, "2-D Shape Representation and Recognition by Lattice Computing Techniques," *Hybrid Artificial Intelligence Systems*, pp. 391–398, 2010.

[50] V. G. Kaburlasos and S. E. Papadakis, "A granular extension of the fuzzy-ARTMAP (FAM) neural classifier based on fuzzy lattice reasoning (FLR)," *Neurocomputing*, vol. 72, no. 10-12, pp. 2067–2078, 2009.

[51] S. E. Papadakis and V. G. Kaburlasos, "Piecewise-linear approximation of nonlinear models based on probabilistically/possibilistically interpreted intervals' numbers (INs)," *Information Sciences*, vol. 180, no. 24, pp. 5060–5076, 2010.