

Vision Based Adaptation to Kernelized Synergies for Human Inspired Robotic Manipulation

Sunny Katyara^{1,2}, Fanny Ficuciello², *Senior Member, IEEE*, Fei Chen^{1,3}, *Senior Member, IEEE*,
Bruno Siciliano², *Fellow, IEEE*, Darwin G. Caldwell¹, *Senior Member, IEEE*,

Abstract—Humans in contrast to robots are excellent in performing fine manipulation tasks owing to their remarkable dexterity and sensorimotor organization. Enabling robots to acquire such capabilities, necessitates a framework that not only replicates the human behaviour but also integrates the multi-sensory information for autonomous object interaction. To address such limitations, this research proposes to augment the previously developed kernelized synergies framework with visual perception to automatically adapt to the unknown objects. The kernelized synergies, inspired from humans, retain the same reduced subspace for object grasping and manipulation. To detect object in the scene, a simplified perception pipeline is used that leverages the RANSAC algorithm with Euclidean clustering and SVM for object segmentation and recognition respectively. Further, the comparative analysis of kernelized synergies with other state of art approaches is made to confirm their flexibility and effectiveness on the robotic manipulation tasks. The experiments conducted on the robot hand confirm the robustness of modified kernelized synergies framework against the uncertainties related to the perception of environment.

I. INTRODUCTION

Robotic manipulation refers to the ways a robot perceives and interacts with the objects in the environment. The knowledge of the environment i.e, the type and location of objects in the world is a deterministic aspect to improve the performance of robots to manipulate, **similar to humans**. Building robot hands that emulate the functionalities of humans is challenging owing to the limitations of existing software and hardware [1]. However, few attempts have been made to mimic the design and control of human hand for flexible grasping and manipulation [2][3][4] but are still lagging behind the human capabilities. The control architecture of humans reveals that they exploit reduced subspace, called postural synergies [5] [6] to plan and regulate their whole body movements. Inspired from postural synergies, a framework called **kernelized synergies** has been developed that reuses the same subspace for robust grasping and dexterous manipulation [7].

The kernelized synergies framework however assumes the pose of objects to be known for grasp execution. But the human inspired robotic manipulation requires to recognize

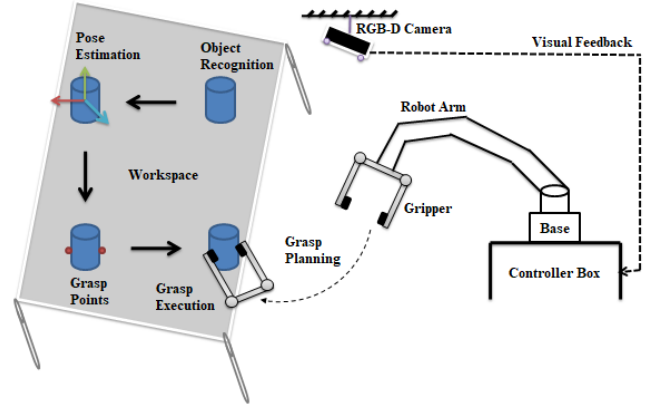


Fig. 1: Sketch of vision based adaptation to robotic manipulation. Camera detects the object in the scene and its pose is estimated which is given to controller box for grasp planning and execution.

and estimate the object pose in run time using **visual perception** as illustrated in Fig. 1. The four key steps to the *vision based robotic manipulation* are (1) bringing gripper within the desirable work-space (2) recognizing and estimating the object pose (3) planning a suitable candidate grasp (4) executing the grasping and manipulation actions. The kernelized synergies framework however considers all the major steps except the object recognition and pose estimation.

The object recognition and pose estimation techniques are still evolving and the algorithms ranging from object semantic segmentation to end-end learning have been exercised for various robotic tasks [8][9][10]. A model based object recognition framework is proposed in [11] for robotic manipulation. The framework is robust against scalability, complexity and latency of the scene but requires multi-view instances for object recognition. A modified object detection approach based on single image segmentation using discriminative classifier is suggested in [12] for robotic grasping. Such technique requires a large data base for object reconstruction and does not work well in the cluttered environments. To alleviate the need of building data base and performing well in the complex scenarios, an object detection framework is discussed in [13] for robotic grasping. The framework however uses model approximations on the filtered point cloud but does not achieve desired accuracy since only two models i.e, spherical and cylindrical are considered for approximation. In our research, we update this pipeline by mitigating the need of using two specific RANSAC models and proposes to exploit additionally a Sup-

This research is supported by projects “LEARN-REAL” funded by EU H2020 ERA-Net Chist-Era and “HARMONY” funded by EU H2020 R&I under grant agreement No 101017008. (Corresponding author: Fei Chen)

¹ Department of Advanced Robotics, Istituto Italiano di Tecnologia, Via Morego 30, 16163, Genova, Italy (e-mail: name.surname@iit.it).

² Department of Information Technology and Electrical Engineering and PRISMA Lab, University of Naples Federico II, Naples 80125, Italy (e-mail: name.surname@unina.it).

³ Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Hong Kong (e-mail: f.chen@ieee.org).

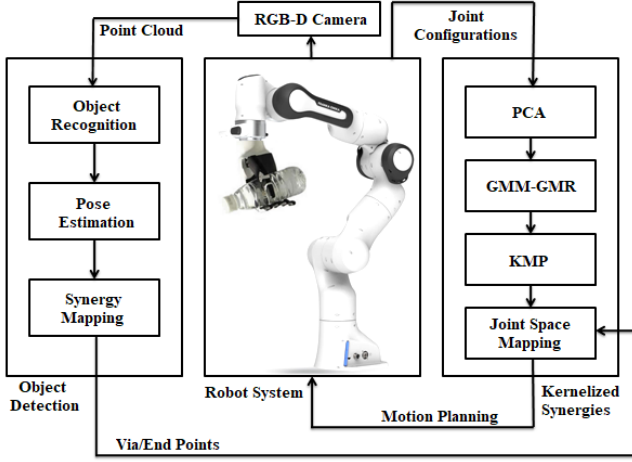


Fig. 2: Block diagram of updated kernelized synergies framework.

port Vector Machine (SVM) classifier for object recognition.

With reference to [7], the key contributions of this work are (1) to upgrade the kernelized synergies framework with visual perception for autonomous grasping and manipulation (2) to compare the performance of kernelized synergies framework with other state of art methods using two metrics i.e, normalized square error and primitive accuracy (3) to apply modified framework to perform three different daily life tasks i.e, mounting bulb on the socket, squeezing lemon into the water, and spraying cleanser on the board.

II. RESEARCH METHODOLOGY

Figure 2 shows the block diagram of research methodology adopted to perform the *human inspired robotic manipulation* tasks. The methodology starts with teaching basic grasping and manipulation primitives to the robot hand on the given training objects in Fig. 3 by tele-operating it and recording the corresponding joint hand configurations. The Principle Component Analysis (PCA) is applied to all the individual tasks and the directions of highest data variance are selected as the predominant postural synergies. The postural synergies evolve over the duration of demonstration to obtain the corresponding synergistic trajectories. The given synergistic trajectories are approximated with Gaussian components using Gaussian Mixture Model (GMM) and the Gaussian Mixture Regression (GMR) generates the reference trajectory to reproduce the taught grasping and manipulation primitives. In order to generalize the learned synergy subspace to wider set of objects, the idea of Kernelized Movement Primitives (KMP) is exploited to deal with the environmental descriptors i.e, via-points and end-points for the new object's shape and size respectively. The kernel trick used in KMP preserves the probabilistic properties of synergy subspace such that it can be reused for grasping and manipulation of unknown objects. The geometrical features of the new objects are estimated from their point cloud using RANSAC algorithm with Euclidean clustering for object segmentation and the SVM classifier, trained on objects in Fig. 3, is used for object recognition. The pose of object is enumerated locally from the centroid of recognized object's



Fig. 3: Training data set, (a) ball, (b) biscuit (c) bottle (d) bowl, (e) box, (f) bulb, (g) coke can, (h) cone, (i) cream, (j) smart card, (k) cube, (l) disk, (m) duster, (n) eraser, (o) glue, (p) highlighter, (q) juice, (r) lemon, (s) mobile, (t) mouse, (u) screw-driver, (v) soap (w) spray, (x) sticky-notes .

point cloud. The estimated pose of object is then transformed into respective synergistic values, which then act as environmental descriptors for the KMP to update its reference trajectory to adapt to the targeted object accordingly. Finally, the mapping from synergistic subspace to the joint space helps in performing the given tasks on the robot system.

A. Kernelized Synergies Framework

The robot hand shown in Fig. 4 (a) is tele-operated on training objects in Fig. 3 and the mean hand positions $\hat{\theta}_k = \theta_k - q_0$, with θ_k and q_0 representing the current hand configuration and nominal hand posture respectively, are recorded. The mean hand positions are then concatenated into a row vector to construct a configuration matrix $C = [\hat{\theta}_1, \dots, \hat{\theta}_K]^T$ and the PCA is applied on C to obtain the lower dimensional representation, called synergistic subspace $\hat{E}$ [14]. Where, $\hat{E}$ is called synergy matrix, which numerically characterizes the synergistic subspace and is accessed by properly selecting the appropriate values of synergy co-efficients (PCA column vector) i.e, $e_i = e_{g1} + e_{m1}, e_{g2} + e_{m2}, \dots, e_{gn} + e_{mn}$ for a given object to be grasped (g) and manipulated (m) and are determined by Eq. 1, where $\hat{E}^\dagger$ is the pseudo inverse of synergy matrix.

$$e_i = \hat{E}^\dagger(\theta_i - q_0) \quad (1)$$

The synergistic subspace of two predominant postural components, estimated using training objects in Fig. 3, for the robot hand in Fig. 4 (a), is illustrated in Fig. 4 (b and c). The computed subspace reveals that the first synergistic component regulates the medial and proximal joints of each finger while the second synergistic component controls the relative motion of index finger and thumb during grasping and manipulation. Further, such a relative motion of index finger and thumb helps in switching from one grasping position to other while an object is being manipulated.

The synergistic co-efficients in Eq. 1 for each object in Fig. 3 evolve over the duration (t) of demonstration to obtain the corresponding synergistic trajectories $e(t)$. The synergistic trajectories define the timing profile of synergistic co-efficients for the demonstrations over the given objects in Fig. 3. The synergistic trajectories are then approximated with Gaussian components $e(t) \sim \sum_{n=1}^N \pi_n \mathcal{N}(\mu_n, \Sigma_n)$ using GMM, with π_n, Σ_n, μ_n representing the prior probability, covariance and mean of the n^{th} Gaussian respectively and N being the total number of Gaussian components used. Further, a reference synergistic trajectory, to be followed by the robot hand to reproduce the taught postures, is generated

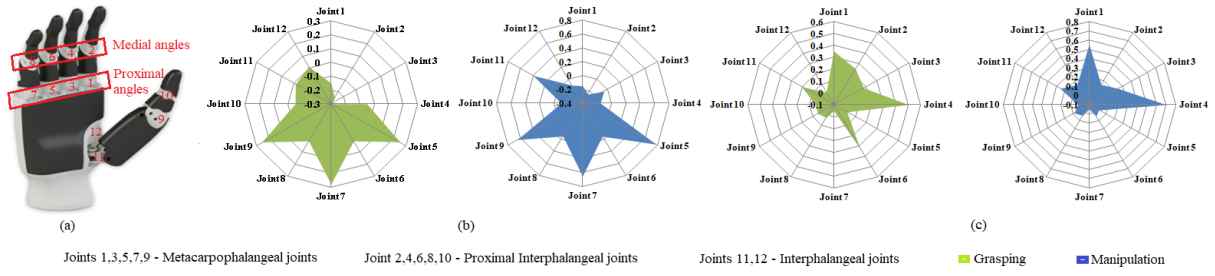


Fig. 4: Graphical representation of computed synergistic subspace of INSPIRE robot hand, (a) represents the 6 DOF robot hand, (b) and (c) are the first two predominant synergies with respective grasping (e_g) and manipulation (e_m) components.

by conditioning the joint probability distribution of the GMM i.e., $e_n | t \sim \mathcal{N}(\mu_n, \Sigma_n)$ using GMR.

To generalize the learned synergy subspace and to adapt to the new objects, the idea of KMP is exploited. For a new instance t^* (time sampling of new object), the expected mean and co-variance of the synergistic trajectory are computed using Eq. 2 [15], with k^* being a kernel function i.e., $k^* = [k(t^*, t_1), k(t^*, t_2), \dots, k(t^*, t_N)]$, K is the corresponding kernel matrix and λ is a regularization parameter.

$$\begin{aligned} \mathbb{E}(e(t^*)) &= k^*(K + \lambda I)^{-1} \mu \\ \mathbb{D}(e(t^*)) &= \frac{N}{\lambda} (k(t^*, t^*) - k^*(K + \lambda \Sigma)^{-1} k^{*T}) \end{aligned} \quad (2)$$

The priorities (Υ) (i.e., weightage to task execution) can be assigned to different interaction tasks in the kernelized synergies framework and are formulated by Eq. 3 [15], which represents the product of R Gaussian distributions in the kernelized synergistic subspace with $r = 1, 2, \dots, R$.

$$\mathcal{N}(\mu_n^T, \Sigma_n^T) \propto \prod_{r=1}^R \mathcal{N}(\hat{\mu}_{n,r}, \hat{\Sigma}_{n,r} / \Upsilon_{n,r}) \quad (3)$$

The compliance in the kernelized synergies for flexible interactions with the objects is introduced by the model of soft synergies $\Delta q_{ref} = \hat{E} \Delta e$ with $\Delta q = \Delta q_{ref} - C_h \Delta \tau$, where C_h represents the robot hand compliance matrix and τ is the vector of robot hand joint torques. The soft synergies, defining adaptive lower dimensional subspace, are useful for regulating the contact forces for soft object interaction using synergistic co-efficients. Therefore, the grasp established with kernelized synergies can be now defined by Eq. 4 in terms of forces balancing the grasped object.

$$f_c = G^\dagger \omega + \xi \Delta e \quad (4)$$

Where, f_c represents the vector of contact forces between the object and robot hand fingers, $G^\dagger$ is the pseudo-inverse of grasp matrix (i.e., defining kinematic and dynamic characteristics of robot hand), ω is the vector of external forces and wrenches applied to the grasped object, ξ is the subspace of internal forces and Δe is the change in kernelized synergistic co-efficients. The kernelized synergies co-efficients interpolated using Eq. 2 are modulated in Eq. 4 until the steady state grasping posture is achieved and it is conditioned by the threshold on the robot hand motors' current (being a feedback signal) to ensure the grasp stability.

B. Object Detection and Synergy Mapping

Algorithm 1: Object Detection & Synergy Mapping

```

input :  $P_c \rightarrow$  point cloud of scene
output:  $e_O \rightarrow$  synergistic values of object
while ( $p^{(i)} \in P_c \rightarrow P_p$ ) do
     $function(P_p) \rightarrow (Inliers, Outliers)$ 
     $P_p := Outliers$ 
     $function(P_p, C) \rightarrow (O);$ 
    if  $d(x, y) \leq \epsilon | MinPts$  then
         $O;$ 
    else
         $\phi \rightarrow$  discard point;
    end
end
foreach ( $shape, color \in O$ ) do
     $function(l, \psi) = SVM(O, model);$ 
    if ( $shape, color \cap (model) \rightarrow object\ detected$ ) then
         $return(l, \psi);$ 
    else
         $\phi \rightarrow$  object not found;
    end
     $P_o^c = centroid(\psi)$ 
     $P_o^b = transformation(P_o^c)$ 
     $e_O = mapping(P_o^b)$ 
end

```

The point cloud of the scene, consisted of objects placed on the table, is acquired using Intel RealSense D435 camera and is shown in Fig. 5 (a). The raw point cloud P_c is filtered to remove the adversarial data points and is down-sampled to reduce the processing time. Further, RANSAC algorithm is used to separate the large planar component representing the table and the remaining points are grouped into different clusters that represent the candidate objects in the processed point cloud P_p in Fig. 5 (b). The essential points in P_p , expressed in k-d tree, are grouped according to Euclidean clustering criteria, defined by Eq. 5.

$$O = \int_{i=0}^C N_\epsilon(x_i | MinPts) : \{y_i | d(x_i, y_i) \leq \epsilon\} dx \quad (5)$$

Where, ϵ represent the radius of neighbour N_ϵ of a point, $MinPts$ is the minimum number of neighbouring points

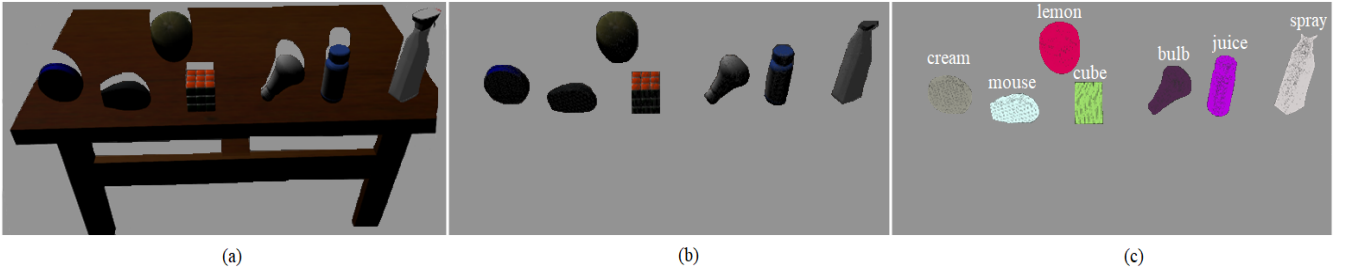


Fig. 5: Object segmentation and recognition, (a) represents the raw point cloud of the scene, (b) illustrates the processed point cloud with candidate objects (dense clusters) after filtering out the adversarial data points and removing the larger planar component (table) using RANSAC algorithm, (c) indicates the candidate objects being recognized and labeled by the trained SVM classifier.

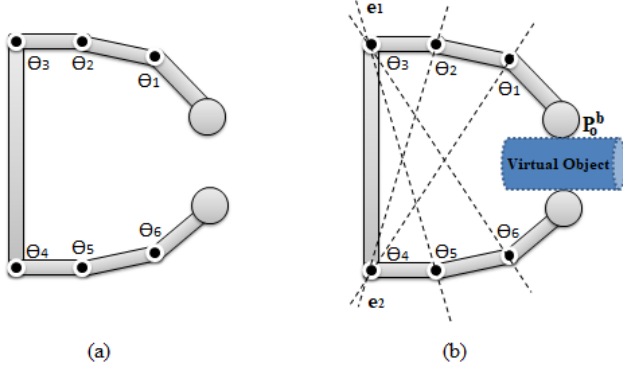


Fig. 6: Object based synergistic mapping to robot hand. (a) represents the robot hand without synchronous (synergistic) control, (b) illustrates the robot hand with virtual object similar to real one at the contact points for determining the corresponding synergistic components. The dashed lines represent the coordinated motion of respective robot hand joints within the synergistic subspace.

within ϵ , point y can be grouped to point x , if the distance between them is less than or equal to ϵ otherwise discarded and C is the number of clusters in the processed point cloud.

The candidate objects found in term of clusters using Eq. 5 are processed further to capture their shape and color characteristics using respective (shape and color) histogram features. Having segmented objects, the SVM classifier trained on objects in Fig. 3 over 10 different orientations of each (for generalization to distinct daily life objects with similar shapes), it successfully recognizes (ψ) and labels (l) the given objects in Fig. 5 (c). The performance of SVM classifier is assessed using confusion matrix in Fig. 9 as an evaluation metric. It is evident from Fig. 9 that the trained classifier has 84% accuracy over 240 instances in the data set.

Such a approach however ensures to grasp and manipulate novel objects but the approximations made with given perception pipeline may introduce some errors in the object recognition i.e., sparse point cloud, occluded object, variational features etc. Thanks to the probabilistic nature of kernelized synergies framework that uses Gaussian kernel having infinite Hilbert-Space, the system is robust against such perceptual errors. Finally, the pose of detected object is calculated locally by averaging all the data in the point cloud to determine the 3D dimensions of their centroid (i.e., defining respective reference frame for pose information).

The extracted geometrical features are in local camera frame and need to be transformed into robot base frame for

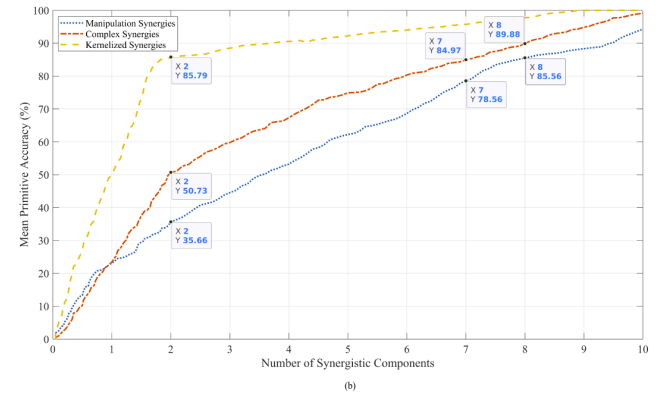
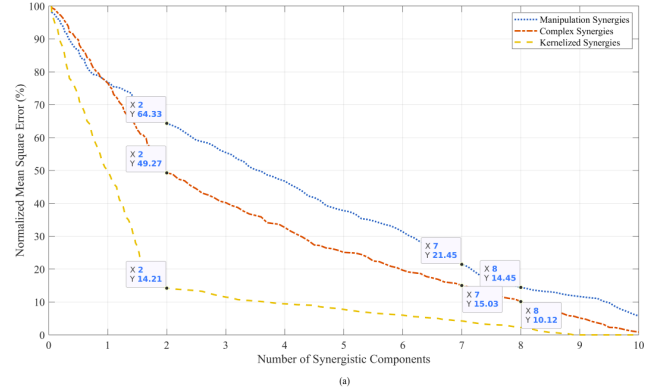


Fig. 7: Performance evaluation of kernelized synergies against two different state of art techniques. Kernelized synergies achieve desired results on normalized square error (a) and primitive accuracy (b) by exploiting the reduced number of synergistic components as compared to manipulation and complex synergies approaches.

precise targeted grasping and manipulation. The object pose in robot base frame is determined by Eq. 6, where T_c^{mr} is the transformation between camera frame and marker, T_{mr}^b is a transformation between marker and base frame, P_o^c and P_o^b are the object poses in camera and base frames respectively.

$$P_o^b = T_{mr}^b T_c^{mr} P_o^c \quad (6)$$

Further, the pose of object needs to be transformed into corresponding synergistic values to update the via/end points of KMP in Eq. 2. For this, a virtual object of dimensions similar to the targeted shape is defined in robot hand Cartesian space in Fig. 6. The relationship between robot hand

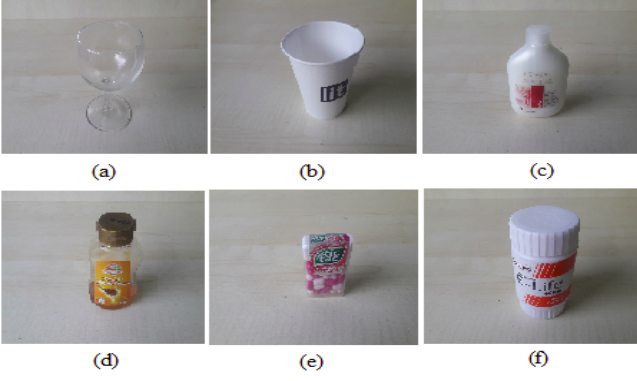


Fig. 8: Set of test objects, (a) glass, (b) cup, (c) conditioner, (d) honey, (e) tic-tac, (f) food-supplement.

Cartesian space and synergistic subspace is given by Eq. 7.

$$\dot{e}_O = A_{mo}^\dagger C_h \hat{E}^\dagger \dot{P}_O^b \quad (7)$$

Where, $\dot{P}_O^b$ is the robot hand Cartesian space velocity due to the object, C_h is the hand compliance matrix, A_{mo} is the motion transfer matrix and † represents the pseudo inverse of respective quantities. The Euler integration at sampling time of 1 ms is applied to Eq. 7 to get the synergistic values of via/end points. Finally, the mapping from synergistic subspace to joint space is required to execute the given task on the robot hand and is defined to be $\theta = \hat{E}e + \theta_0$, where θ_0 represents the initial robot hand configuration.

The complete pipeline of object detection and synergy mapping is summarized in Algorithm 1.

III. COMPARATIVE ANALYSIS

To better understand the performance and potential of kernelized synergies framework, it is compared with other state of art techniques such as; manipulation synergies [16] and complex synergies [17] using two different evaluation metrics i.e., normalized square error (NSE) and primitive accuracy (PA) in Fig. 7. The robot hand trained on objects in Fig. 3, is evaluated on test objects in Fig. 8. The robot hand performs grasping (precision) and manipulation (rotation) of test objects and the difference between robot hand configurations defined by human subject(ground-truth) and established by considered techniques is computed to define the given metrics in Fig. 7. The values of NSE and PA plotted in Fig. 7, are the average of all the hand configurations established over test objects in Fig. 8. Note that for simplicity, all the objects are placed in orthogonal direction to robot hand palm and their grasping postures are pre-defined.

As shown in Fig. 7, the manipulation synergies have high NSE and low PA because they are task specific and for any unknown object it requires to recompute the entire synergistic subspace on the new data set. Therefore, it completely changes the shape of whole synergy subspace and in particular the grasping properties of predominant postural synergies are not preserved. Such framework achieves the desired performance i.e., 15% NSE and 85% PA by exploiting almost 8 synergistic components, which itself violates the motivation behind using synergies i.e., bringing simplicity.

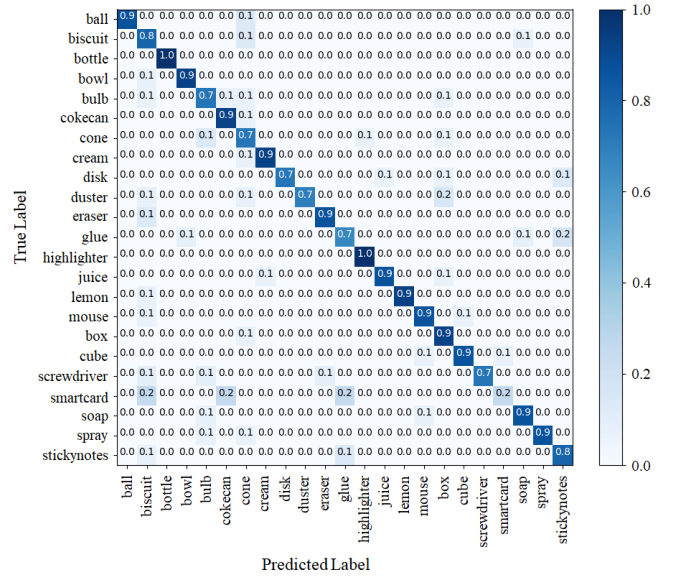


Fig. 9: Normalized confusion matrix of trained SVM classifier.

However, the complex synergies have relatively low NSE and high PA as compared to manipulation synergies and attain desired accuracy but again at the cost of large number of synergistic components (less than manipulation synergies) to reproduce the required hand configurations. Although, this technique preserves the grasping properties of synergistic subspace and the manipulation tasks are executed by the additional synergistic components that are used to correct the corresponding hand configuration with respect to grasping posture. Although, it does not require to re-compute the synergy subspace for every new entry in contrast to manipulation synergies but it increases the size of synergy subspace and is computationally inefficient. Further, the properties of grasping synergies are preserved locally and can not be generalized to even different dimensions of same objects.

It is evident from Fig. 7 that the kernelized synergies have lowest NSE and highest PA as compared to other two techniques mainly due to two reasons; (1) the use of movement primitives to adapt to the dimensions of new geometrical shapes and (2) the application of kernel trick which preserves the probabilistic properties of computed subspace globally such that it can be reused for grasping and manipulation of different objects. Further, this framework achieves desired results by utilizing just first two predominant synergistic components with and without visual perception.

IV. RESULTS AND DISCUSSION

To experimentally evaluate the performance of updated kernelized synergies framework, three different daily life tasks i.e., (1) mounting bulb on the socket, (2) squeezing lemon into the water, and (3) spraying cleanser on the board are performed. Note that the appropriate values of e_m and e_g are determined according to Eq. 7.

In the first task, the robot hand performs bulb mounting task by exploiting rotation primitives of its synergy subspace, similar to humans. For simplicity, the expired bulb has been already removed from the socket and the robot hand is

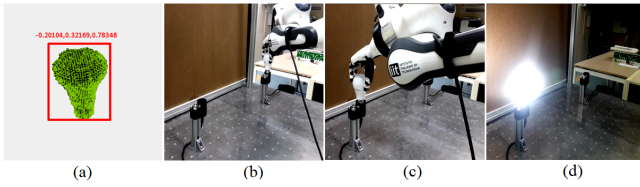


Fig. 10: A robot mounting bulb on the socket, (a) is the estimated pose of bulb in the camera frame, (b) shows the robot hand grasping the bulb in its tripod posture, in (c) robot brings the bulb over the socket, (d) represents the bulb after being screwed into the socket.

asked to mount the new one. Firstly, the eye-to-hand camera estimates the pose of the bulb using Eq. 6 as shown in Fig. 10 (a) and updates the kernelized synergies subspace according to Eq. 7. The updated values of kernelized synergies are computed using Eq. 2 that help in mounting the bulb on the socket. The robot hand grasps the bulb in tripod posture at $e_{g1} = -0.18$, $e_{g2} = 0.21$ in Fig. 10 (b) and places it on the socket using robot arm movements in Fig. 10 (c). The bulb is then rotated clockwise at $e_{m1} = -0.19$ to -0.05 , $e_{m2} = 0.21$ to 0.37 such that it is screwed into the socket as shown in Fig. 10 (d). In this task, the force between the robot hand fingertips and the bulb is modulated approximately from 2.38 N to 3.57 N according to Eq. 4. It is due to the reason that the resistance between the bulb's rings and the socket increases during the final phase when it is about to mount on the socket and thus requires to apply more strength.

During the second task, the robot hand utilizes the translation primitives of its synergy subspace to squeeze the lemon into the water to prepare the lemon juice. With the pose of lemon estimated in Fig. 11 (a), the robot hand grips it in a tripod posture in Fig. 11 (b) at $e_{g1} = 0.12$, $e_{g2} = 0.32$ and brings it over the glass using the robot arm movements in Fig. 11 (c). The robot hand stretches in its index and middle fingers at $e_{m1} = 0.13$ to 0.26 , $e_{m2} = 0.33$ to 0.47 such that the lemon is pressed and its juice is sprinkled into the glass of water in Fig. 11 (d). Alike first task, the interaction force is modulated from 2.38 N to 4.16 N according to Eq. 4.

For the third task, the robot hand takes an advantage of priority characteristics of kernelized synergies framework according to Eq. 3, to regulate its two parts independently. In this task, using pose information of spray bottle in Fig. 12 (a), the robot hand initially assumes the pre-grasping pose in Fig. 12 (b) and then grasps it in Fig. 12 (c) at $e_{g1} = -0.11$, $e_{g2} = 0.38$ with priority $\Upsilon = 0.5$ being assigned to each. The robot hand presses the spray within its tripod posture at $e_{m1} = -0.1$ to 0.14 , $e_{m2} = 0.39$ to 0.48 while the index and ring fingers help in holding the spray in Fig. 12 (d). The action of pressing the spray, results into showering of cleanser on the board. Similarly, in this task, the interaction force between the spray and robot hand is modulated from approximately 2.38 N to 4.76 N according to Eq. 4. Such priority characteristics of kernelized synergies framework are helpful in various multi-digit manipulations such as; holding and typing on smart phone, gripping and pressing the computer mouse, holding receiver and dialing a number on the telephone and many others.

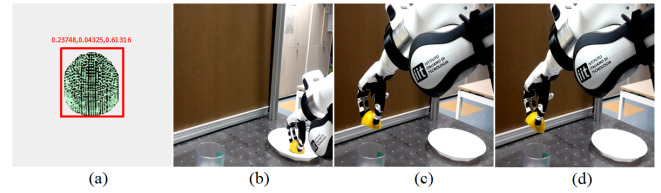


Fig. 11: A robot squeezing lemon into the glass of water, (a) is the estimated pose of lemon, (b) shows the robot hand gripping lemon in its tripod posture, (c) illustrates the robot bringing lemon over the glass, (d) represents the robot hand squeezing lemon into glass.

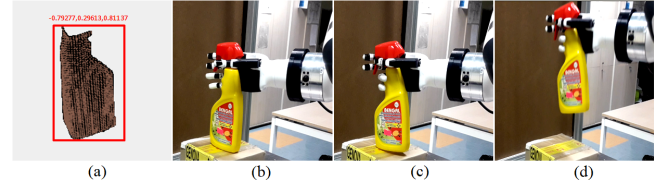


Fig. 12: A robot spraying cleanser on the board, (a) is the estimated pose of spray in the camera frame, (b) is the pre-grasping posture of robot hand, (c) shows the robot hand holding spray with its thumb, little and ring fingers, (d) represents the robot hand spraying against the board with its index and middle fingers.

V. CONCLUSIONS

This research proposed to update the previously developed kernelized synergies framework with visual perception for autonomous grasping and manipulation of novel objects. The point cloud of objects in the scene was extracted with RGB-D camera and was pre-processed and later filtered using RANSAC algorithm to remove the largest planar component in the scene. The remaining points were grouped to represent the candidate objects, based on the Euclidean clustering segmentation criteria. The candidate objects in the filtered point cloud were recognized and labeled using the trained SVM classifier and the pose was estimated locally by computing the centroid (defining 3D reference frame) of each candidate object. The geometrical features of the detected objects were in image plane and the suitable transformations were applied to obtain the desired pose in the robot base frame for task execution. The use of RANSAC algorithm however introduced some approximation inaccuracies in the object recognition and pose estimation pipeline but due to the probabilistic nature of kernelized synergies, it was robust against the perceptual errors. Further, the performance of kernelized synergies framework was compared with the other state of art techniques and it was evident that the kernelized synergies are effective, efficient and robust in performing various robust grasping and fine manipulation tasks. Finally, the updated framework was experimentally evaluated on three different daily life tasks i.e., mounting bulb on the socket, squeezing lemon into the water, and spraying cleanser on the board and the results confirm the task agnostic approach of updated kernelized synergies framework.

For future work, the tactile information will be integrated into the kernelized synergies framework to do visuo-tactile dexterous manipulation. Further, the updated framework can also be extended to anthropomorphic dual arm-hand system for whole body manipulation in the cluttered environments.

REFERENCES

- [1] A. Billard and D. Kragic, "Trends and challenges in robot manipulation," *Science*, vol. 364, no. 6446, 2019.
- [2] M. G. Catalano, G. Grioli, E. Farnioli, A. Serio, C. Piazza, and A. Bicchi, "Adaptive synergies for the design and control of the pisa/iit soft hand," *The International Journal of Robotics Research*, vol. 33, no. 5, pp. 768–782, 2014.
- [3] R. Deimel and O. Brock, "A novel type of compliant and underactuated robotic hand for dexterous grasping," *The International Journal of Robotics Research*, vol. 35, no. 1-3, pp. 161–185, 2016.
- [4] C. Della Santina, C. Piazza, G. Grioli, M. G. Catalano, and A. Bicchi, "Toward dexterous manipulation with augmented adaptive synergies: The pisa/iit soft hand 2," *IEEE Transactions on Robotics*, vol. 34, no. 5, pp. 1141–1156, 2018.
- [5] M. Santello, M. Flanders, and J. F. Soechting, "Postural hand synergies for tool use," *Journal of neuro-science*, vol. 18, no. 23, pp. 10 105–10 115, 1998.
- [6] F. Ficuciello, A. Federico, V. Lippiello, and B. Siciliano, "Synergies evaluation of the schunk s5fh for grasping control," in *Advances in Robot Kinematics 2016*. Springer, 2018, pp. 225–233.
- [7] S. Katyara, F. Ficuciello, D. Caldwell, B. Siciliano, and F. Chen, "Leveraging kernelized synergies on shared subspace for precision grasp and dexterous manipulation," *arXiv preprint arXiv:2008.11574*, 2020.
- [8] A. Collet, D. Berenson, S. S. Srinivasa, and D. Ferguson, "Object recognition and full pose registration from a single image for robotic manipulation," in *2009 IEEE International Conference on Robotics and Automation*. IEEE, 2009, pp. 48–55.
- [9] Y. Yang and Q.-X. Cao, "Monocular vision based 6d object localization for service robot's intelligent grasping," *Computers and Mathematics with Applications*, vol. 64, no. 5, pp. 1235–1241, 2012.
- [10] . Bohg, A. Morales, T. Asfour, and D. Kragic, "Data-driven grasp synthesis—a survey," *IEEE Transactions on Robotics*, vol. 30, no. 2, pp. 289–309, 2013.
- [11] M. Zhu, K. G. Derpanis, Y. Yang, S. Brahmabhatt, M. Zhang, C. Phillips, M. Lecce, and K. Daniilidis, "Single image 3d object detection and pose estimation for grasping," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 3936–3943.
- [12] A. Collet, M. Martinez, and S. S. Srinivasa, "The moped framework: Object recognition and pose estimation for manipulation," *The international journal of robotics research*, vol. 30, no. 10, pp. 1284–1306, 2011.
- [13] F. Ficuciello, A. Miglinozzi, G. Laudante, P. Falco, and B. Siciliano, "Vision-based grasp learning of an anthropomorphic hand-arm system in a synergy-based control framework," *Science Robotics*, vol. 4, no. 26, 2019.
- [14] . Ficuciello, G. Palli, C. Melchiorri, and B. Siciliano, "A model-based strategy for mapping human grasps to robotic hands using synergies," in *2013 IEEE/ASME International Conference on Advanced Intelligent Mechatronics*. IEEE, 2013, pp. 1737–1742.
- [15] Y. Huang, L. Roza, J. Silverio, and D. G. Caldwell, "Kernelized movement primitives," *The International Journal of Robotics Research*, vol. 38, no. 7, pp. 833–852, 2019.
- [16] N. Jarrass e, A. T. Ribeiro, A. Sahbani, W. Bachta, and A. Roby-Brami, "Analysis of hand synergies in healthy subjects during bimanual manipulation of various objects," *Journal of neuro engineering and rehabilitation*, vol. 11, no. 1, p. 113, 2014.
- [17] G. Palli, F. Ficuciello, U. Scarcia, C. Melchiorri, and B. Siciliano, "Experimental evaluation of synergy-based in-hand manipulation," *IFAC Proceedings Volumes*, vol. 47, no. 3, pp. 299–304, 2014.