From Poking to Grasping: Active Shape Completion for Improved Robot Manipulation

Lukas Rustler and Matej Hoffmann

***Abstract*— Robust robot grasping of a priori unknown objects remains challenging when the object is only partially observed. While single-view shape completion methods infer missing geometry, their predictions often contain high uncertainty that leads to grasp failures. Active perception solves this by gathering new targeted information. This extended abstract synthesizes progress using a visuo-tactile framework across three progressive methodologies—Act-VH, VISHAC, and ShapeGrasp. We outline the transition from uncertainty-driven haptic poking to a closed-loop "grasp-and-complete" pipeline, demonstrating how physical interaction and grasp attempts can seamlessly serve as active perception queries to refine implicit shape models and achieve state-of-the-art manipulation success.**

I. Introduction

A core challenge in unstructured robotic manipulation is acquiring complete and accurate 3D object models from partial visual data. Due to self-occlusions, sensor noise, or adversarial object materials, visual-only reconstructions possess high uncertainty, which negatively impacts downstream tasks such as grasping. Active perception as defined by Ruzena Bajcsy [1] overcomes these limitations by deliberately interacting with and exploring the environment—in our case poking or grasping the object to extend the surface for which the model is accurate.

This contribution presents the evolution of our visuo-tactile shape completion pipeline (estimation of the whole shape of an object from partial information). We detail the progression from Act-VH [2], which uses a dedicated 3D printed finger to sample uncertain regions, through VISHAC [3], which improves efficiency using a gripper and free space constraints, and finally to ShapeGrasp [4], which elegantly merges exploration and manipulation by extracting tactile and volumetric constraints directly from grasping attempts.

II. Related Work

Visual Shape Completion. Even though visual-only shape completion is an intuitive solution to the problem, it is often static and lacks the advantages of active perception. The field evolved from using a database of objects [5], primitives [6] or symmetry assumptions [7] to more generalizable methods using Machine Learning (ML). The literature often uses voxel-based reconstructions [8], [9], Signed Distance Function (SDF) [10], [11], [12], and Transformers-based networks [13], [14], [15].

Tactile Shape Completion. Tactile approaches are, on the other hand, almost implicitly active, as random tactile exploration would be unfeasible in the real world. The literature often uses Gaussian Processes (GP) [16] to guide the haptic exploration [17], [18], [19]. Having an object modeled using the Gaussian distribution automatically adds the option to infer the uncertainty of each point. An improved GP formulation was used by Bonzini *et al.* to autonomously discover symmetries that help with exploration [20].

Visuo-Tactile Shape Completion. The combination of visual and tactile input is the closest to how humans approach active perception. An initial information is obtained using vision, and the model (belief) is updated using touch. There are approaches using Gaussian Processs (GPs) [21], [22], [23], [24], Convolutional Neural Networks (CNNs) [25], [26] or other deep learning methods for higher resolution [27], [28]. Our works [2], [3], [4] also belong here.

III. Implicit Visuo-Tactile Representation

All of our frameworks share the core representation of the object O as an implicit surface defined as

$$O = \{\mathbf{x} \in \mathbb{R}^3 \mid \ f(\mathbf{x};\boldsymbol{\theta},\mathbf{z}) = 0\}, \tag{1}$$

where f is a shape completion network with parameters $\boldsymbol{\theta}$ and $\mathbf{z}$ is the latent vector optimized using the points $\mathbf{x}$ of the current point cloud $\mathcal{X}$. The function f is based on the work of Gropp *et al.* [12] and is represented by an Multi-Layer Perceptron (MLP) that learns an SDF, i.e., learns how to estimate the signed distance of each point from an underlying surface. The input point cloud is created using the fusion of multiple senses. The given input is different for each of the works we present, but can be any of the following:

- **Visual Points:** Camera point cloud.
- **Tactile Points:** Points extracted from tactile exploration actions—either poking or grasping.
- **Free Space:** Space that must be empty in object-centric representation—either space explored by the robot or space where the body of a gripper is located.

IV. Evolution of Active Perception Strategies

We progressively changed the type of active exploration used in our pipelines. We started with a dedicated poking finger attached to a robotic manipulator, continued with using a closed gripper as a poking device, and ended with a closed-loop grasp-and-complete approach.

Lukas Rustler and Matej Hoffmann are with the Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague, matej.hoffmann@fel.cvut.cz, lukas.rustler@fel.cvut.cz. We thank Jens Lundell, Jiri Matas, Ville Kyrki, and Jan K. Behrens for the help as co-authors of the individual papers.

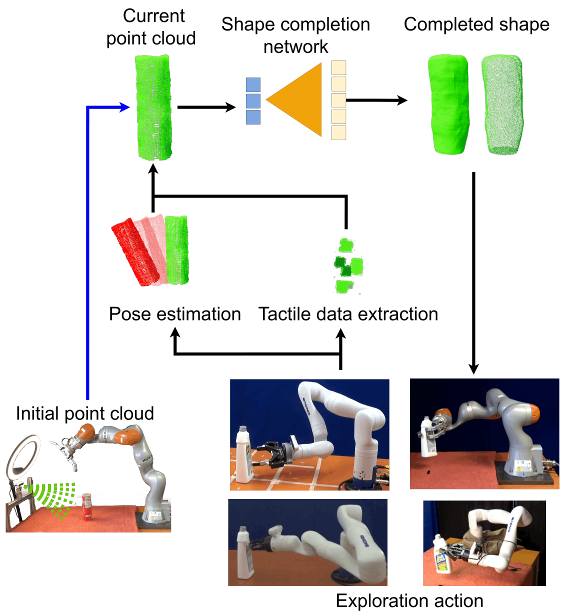

Fig. 1. General schema of the pipeline. An initial point cloud is obtained and used for shape completion. An exploration action (poking or grasping) is performed. Tactile data are captured and current pose (not in Act-VH) is computed. The operation repeats until end condition—maximum steps or successful grasp.

A. Exploration by Poking (Act-VH & VISHAC)

We started with a thought: "If humans are not sure about something using only vision, they go and explore it with their fingers." In the initial work, Act-VH [2], we took it literally and put a finger on a robotic manipulator (see Fig. 1). We also wanted to explore the most uncertain areas. The first uncertainty measure was rather empirical. The shape completion is iterative, and we observed that in the later iterations some parts are almost not changing, and some are changing every iteration. We thus assumed that the points that have high variance between iterations are uncertain from the point of view of the network.

The second work, VISHAC [3], addressed the limitations of the previous pipeline. We removed the additional finger and used closed jaws of a gripper as the poking device. Then, we employed a more mathematically grounded uncertainty metric. Takashi [29] proved that $f(\mathbf{x})$ that meets the condition of the Eikonal Equation $\|\nabla_{\mathbf{x}} f(\mathbf{x})\| = 1$ on a Riemann manifold M is an SDF to a hypersurface M. We used this fact to utilize the loss $(\|\nabla_{\mathbf{x}} f(\mathbf{x};\boldsymbol{\theta},\mathbf{z})\| - 1)^2$ as a measure of uncertainty for each point $\mathbf{x}$. In addition, VISHAC solved other minor issues, such as pose estimation so that objects can move freely on a table, reduced the time needed for a touch by 50%, and added the option of having multiple objects on a table at the same time.

B. Exploration by Grasping (ShapeGrasp)

However, humans usually do not poke objects to infer their properties. More often we just look and attempt to grasp. We tackled this in our last work, ShapeGrasp [4]. The pipeline removes the distinction between exploration and execution. Instead of dedicated poking, the system attempts a full grasp based on the initial visual completion. The grasp is planned using a rigid-body physics simulation that evaluates vertical slippage, rotation slippage, and contact maintenance.

If the real-world grasp attempt fails, the failure serves as an informative active perception query. The tactile sensors provide evidence of the true surface, while the gripper's body provides dense free space constraints—the space where the gripper is in the real world must be empty in the implicit surface definition of the object. The object's pose is estimated, the new constraints are fused, the implicit shape is refined, and a new grasp is generated.

V. Experimental Highlights

The methodologies were extensively evaluated in simulation and on real-world robotic setups, including a Kinova Gen3 arm with a Robotiq 2F-85 gripper and a KUKA LBR iiwa with a Barrett Hand. The general schema of operation can be seen in Fig. 1. The results for VISHAC and ShapeGrasp are shown in Tab. I.

- **Shape Completion Quality:** Adding active tactile feedback significantly improves geometric accuracy over visual-only baselines. In Act-VH and VISHAC, the touches consistently improve the comparison metrics—Jaccard similarity (JS) and Chamfer distance (CD). In ShapeGrasp, just getting the information as a byproduct of grasping improves the metrics as well.
- **Grasp Success Rate (GSR):** Act-VH demonstrated that 5 uncertainty-driven exploratory touches increased grasp success from 38% to 80% using a real-world two-finger gripper. VISHAC improved the GSR up to 85% after 10 touches (while requiring a similar time). ShapeGrasp further elevated these results by achieving an overall success rate of 91% with the two-finger gripper and 84% with the three-finger gripper. By coupling exploration directly with grasping, ShapeGrasp drastically reduced the time required to achieve a successful grasp compared to the hundreds of seconds required for dedicated haptic exploration.

TABLE I

Comparison of VISHAC and ShapeGrasp using the same object set. The values are computed over 3 repetitions for VISHAC and 10 repetitions for ShapeGrasp.

	VISHAC [3]		**ShapeGrasp** [4]
	5 touches	*10 touches*	
JS (%) ↑	60.07 ± 4.50	63.17 ± 5.91	58.14 ± 6.61
CD (mm) ↓	18.65 ± 3.64	17.35 ± 3.91	18.64 ± 2.21
GSR (%) ↑	82	85	91

VI. Conclusion

The trajectory from Act-VH to ShapeGrasp underscores a powerful paradigm for the active perception community: action and perception are most effective when deeply intertwined. By exploiting physical grasp attempts as active sensory queries, robots can dynamically refine their internal geometric models while actively pursuing manipulation objectives. This unified visuo-tactile framework demonstrates a robust and scalable path toward the autonomous manipulation of unknown objects in unstructured environments.

REFERENCES

[1] R. Bajcsy, "Active perception," *Proceedings of the IEEE*, vol. 76, no. 8, pp. 966–1005, 1988.

[2] L. Rustler, J. Lundell, J. K. Behrens, V. Kyrki, and M. Hoffmann, "Active Visuo-Haptic Object Shape Completion," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 5254–5261, Apr. 2022.

[3] L. Rustler, J. Matas, and M. Hoffmann, "Efficient Visuo-Haptic Object Shape Completion for Robot Manipulation," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2023, pp. 3121–3128.

[4] L. Rustler and M. Hoffmann, "Shapegrasp: Visuo-tactile shape completion while grasping for improved robot manipulation," *To be published.*, 2026.

[5] M. Pauly, N. J. Mitra, J. Giesen, M. H. Gross, and L. J. Guibas, "Example-based 3D Scan Completion," in *Symposium on Geometry Processing*, no. CONF, 2005, pp. 23–32.

[6] R. Schnabel, P. Degener, and R. Klein, "Completion and Reconstruction with Primitive Shapes," in *Computer Graphics Forum*, vol. 28, no. 2. Wiley Online Library, 2009, pp. 503–512.

[7] J. Bohg, M. Johnson-Roberson, B. León, J. Felip, X. Gratal, N. Bergström, D. Kragic, and A. Morales, "Mind the gap-robotic grasping under incomplete observation," in *2011 IEEE International Conference on Robotics and Automation*. IEEE, 2011, pp. 686–693.

[8] J. Varley, C. DeChant, A. Richardson, J. Ruales, and P. Allen, "Shape Completion Enabled Robotic Grasping," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sep. 2017, pp. 2442–2447.

[9] J. Lundell, F. Verdoja, and V. Kyrki, "Robust Grasp Planning Over Uncertain Shape Completions," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Nov. 2019, pp. 1526–1532.

[10] J. J. Park, P. Florence, J. Straub, R. Newcombe, and S. Lovegrove, "DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Jun. 2019, pp. 165–174.

[11] M. Atzmon and Y. Lipman, "SAL: Sign Agnostic Learning of Shapes From Raw Data," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Jun. 2020, pp. 2562–2571.

[12] A. Gropp, L. Yariv, N. Haim, M. Atzmon, and Y. Lipman, "Implicit Geometric Regularization for Learning Shapes," in *International Conference on Machine Learning*. PMLR, Nov. 2020, pp. 3789–3799.

[13] A. Rosasco, S. Berti, F. Bottarel, M. Colledanchise, and L. Natale, "Towards Confidence-guided Shape Completion for Robotic Applications," in *2022 IEEE-RAS 21st International Conference on Humanoid Robots (Humanoids)*, 2022, pp. 580–586.

[14] X. Yan, L. Lin, N. J. Mitra, D. Lischinski, D. Cohen-Or, and H. Huang, "ShapeFormer: Transformer-Based Shape Completion via Sparse Representation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 6239–6249.

[15] S. S. Mohammadi, N. F. Duarte, D. Dimou, Y. Wang, M. Taiana, P. Morerio, A. Dehban, P. Moreno, A. Bernardino, A. Del Bue, and J. Santos-Victor, "3DSGrasp: 3D Shape-Completion for Robotic Grasp," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, May 2023, pp. 3815–3822.

[16] O. Williams and A. Fitzgibbon, "Gaussian process implicit surfaces," in *Gaussian Processes in Practice*, 2006.

[17] Z. Yi, R. Calandra, F. Veiga, H. van Hoof, T. Hermans, Y. Zhang, and J. Peters, "Active Tactile Object Exploration with Gaussian Processes," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct. 2016, pp. 4925–4930.

[18] D. Driess, P. Englert, and M. Toussaint, "Active Learning with Query Paths for Tactile Object Shape Exploration," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sep. 2017, pp. 65–72.

[19] S. Dragiev, M. Toussaint, and M. Gienger, "Uncertainty aware grasping and tactile exploration," in *2013 IEEE International Conference on Robotics and Automation*, 2013, pp. 113–119.

[20] A. A. Bonzini, L. Seminara, S. Macciò, A. Carfì, and L. Jamone, "Robotic Haptic Exploration of Object Shape With Autonomous Symmetry Detection," *IEEE Transactions on Robotics*, vol. 41, 2025.

[21] G. Z. Gandler, C. H. Ek, M. Björkman, R. Stolkin, and Y. Bekiroglu, "Object Shape Estimation and Modeling, Based on Sparse Gaussian Process Implicit Surfaces, Combining Visual Data and Tactile Exploration," *Robotics and Autonomous Systems*, vol. 126, p. 103433, 2020.

[22] S. Ottenhaus, D. Renninghoff, R. Grimm, F. Ferreira, and T. Asfour, "Visuo-Haptic Grasping of Unknown Objects Based on Gaussian Process Implicit Surfaces and Deep Learning," in *2019 IEEE-RAS 19th International Conference on Humanoid Robots (Humanoids)*, Oct. 2019, pp. 402–409.

[23] M. Björkman, Y. Bekiroglu, V. Högman, and D. Kragic, "Enhancing Visual Perception of Shape through Tactile Glances," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Nov. 2013, pp. 3180–3186.

[24] S. Suresh, Z. Si, J. G. Mangelson, W. Yuan, and M. Kaess, "ShapeMap 3-D: Efficient shape mapping through dense touch and vision," in *2022 International Conference on Robotics and Automation (ICRA)*, 2022, pp. 7073–7080.

[25] D. Watkins-Valls, J. Varley, and P. Allen, "Multi-Modal Geometric Learning for Grasping and Manipulation," in *2019 International Conference on Robotics and Automation (ICRA)*, May 2019, pp. 7339–7345.

[26] S. Wang, J. Wu, X. Sun, W. Yuan, W. T. Freeman, J. B. Tenenbaum, and E. H. Adelson, "3D Shape Perception from Monocular Vision, Touch, and Shape Priors," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct. 2018, pp. 1606–1613.

[27] P. K. Murali, C. Wang, D. Lee, R. Dahiya, and M. Kaboli, "Deep Active Cross-Modal Visuo-Tactile Transfer Learning for Robotic Object Recognition," *IEEE Robotics and Automation Letters*, vol. 7, pp. 9557–9564, 10 2022.

[28] E. Smith, D. Meger, L. Pineda, R. Calandra, J. Malik, A. Romero Soriano, and M. Drozdzal, "Active 3D Shape Reconstruction from Vision and Touch," *Advances in Neural Information Processing Systems*, vol. 34, 2021.

[29] T. Sakai, "On Riemannian manifolds admitting a function whose gradient is of constant norm," *Kodai Mathematical Journal*, vol. 19, no. 1, pp. 39 – 51, 1996.