2026 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM) | 979-8-3195-3611-2/26/$31.00 ©2026 EU | DOI: 10.1109/AIM65483.2026.11658209

SESTOSENSO: Proxy-Tactile Perception for Intelligent and Safe Human–Robot Interaction

Simone Borelli*, Francesco Giovinazzo*, Francesco Grella*
Giammarco Caroleo†, Alessandro Albini†, Perla Maiolino†, Giorgio Cannata*
* DIBRIS, University of Genoa, Genoa, Italy
† Oxford Robotics Institute, University of Oxford, Oxford, UK

***Abstract*—Next-generation collaborative robots require enhanced perception and physical awareness to interact safely with humans and operate effectively in unstructured and cluttered environments. This paper presents SESTOSENSO, a framework for proxy-tactile perception and control that combines large-area multimodal robot skin, distributed embedded sensing, middleware layer, perception modules, and reactive control strategies for safe human–robot collaboration. The framework is built around ProxySKIN, which integrates tactile and proximity sensing over the robot body to provide a continuous perceptual capability, from pre-contact awareness to physical interaction. On top of this sensing layer, SESTOSENSO supports key enabling functionalities including proxy-tactile control, safe human–robot interaction, environment perception, and robot self-awareness. This paper presents the overall system architecture and the role of its main components within a unified proxy-tactile framework. The approach is validated in an automotive collaborative scenario, in which the sensorized robotic system assists a human operator during vehicle assembly operations. The results show the feasibility of the proposed framework for supporting safe human-robot collaboration in industrial settings.**

***Index Terms*—Human–robot interaction, tactile sensing, proximity sensing, collaborative robots, robot perception**

I. Introduction

Collaborative robots are increasingly being deployed in industrial environments where they must operate safely in close proximity with humans. However, current human–robot collaboration systems still rely predominantly on vision-based perception and externally installed safety monitoring solutions, which can suffer from limited robustness in cluttered environments and impose additional setup, calibration, and integration constraints that reduce deployment flexibility [1]–[5].

Alternative solutions have been explored through body-integrated sensing technologies, particularly tactile and proximity sensing systems integrated directly onto the robot body. Early robotic skin approaches focused on distributed tactile sensing for contact localization and force estimation during physical interaction [6], [7]. More recent works investigated capacitive and proximity-based sensing technologies capable of detecting nearby humans and obstacles before contact occurs [8]–[11]. These approaches demonstrated the potential of embodied sensing for improving safety and interaction awareness in collaborative robotics. In parallel, recent surveys on perception technologies for collaborative robots highlighted the increasing relevance of multimodal sensing architectures

This work has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101070310. The views and opinions expressed are those of the authors and do not necessarily reflect those of the European Union. Corresponding author: simone.borelli@edu.unige.it

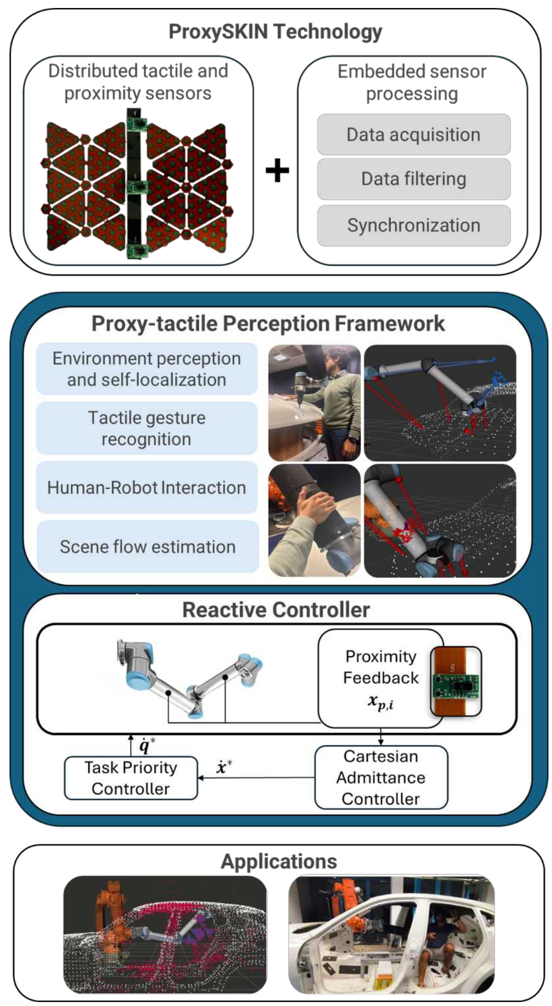

Fig. 1. Conceptual organization of the SESTOSENSO framework into three main layers: (i) the ProxySKIN technology and embedded processing infrastructure, (ii) the proxy-tactile perception and control layer, and (iii) the application layer, where ProxySKIN is validated.

combining tactile, proximity, and proprioceptive information to overcome the limitations of purely vision-based systems, especially in cluttered and unstructured environments [12]. Nevertheless, many existing approaches focus on isolated sensing modalities or specific interaction tasks, while fewer works address the integration of distributed multimodal robot skins within a unified perception and control framework for large-scale collaborative manipulation scenarios.

Within this context, the SESTOSENSO framework integrates a heterogeneous multimodal sensing architecture based on distributed proxy-tactile sensing, embedded processing, perception modules, and reactive control strategies within a unified framework to support robot self-awareness and safe human–robot collaboration. The contribution of this paper is the presentation of the SESTOSENSO framework and its system architecture together with its validation in an automotive collaborative scenario representative of real industrial assembly operations. Detailed algorithmic developments for the individual sensing, perception, and control components are reported in dedicated publications. The main elements of the SESTOSENSO framework include:

- Distributed proxy-tactile sensing technology (ProxySKIN) enabling large-area tactile and proximity sensing over the robot body.
- Perception methods supporting robot self-awareness, environment modeling, and human gestures recognition.
- Control strategies that exploit proxy-tactile feedback to enable safe physical interaction and reactive robot behavior.
- Validation of the proposed framework through an automotive collaborative assembly demonstrator.

The remainder of the paper is organized as follows. Section II introduces the SESTOSENSO project and the ProxySKIN technology enabling distributed tactile and proximity sensing over the robot body. Section III presents the system architecture, including the ProxySKIN sensor network and the embedded processing infrastructure. Section IV describes perception methods that exploit proxy-tactile sensing for environment modeling, human gesture recognition, and robot self-localization. Section V presents control strategies that leverage proxy-tactile feedback to enable safe human–robot interaction and reactive robot motion. Section VI presents the experimental results obtained in an automotive collaborative assembly scenario used to validate the proposed framework. Section VII discusses the experimental findings and the limitations of the proposed approach. Finally, Section VIII concludes the paper.

II. Project Vision

The SESTOSENSO framework builds on the vision of enabling robots to perceive and interact with their surroundings without relying on external sensing infrastructures. Within the SESTOSENSO project [13], this vision is pursued through the development of distributed proxy-tactile sensing technologies that combine proximity and tactile sensors integrated across the robot body, enabling robots to cooperate safely and effectively with humans and adapt to changing operational conditions. The core idea of this project is to exploit miniaturized proximity and tactile sensors distributed across the robot body to gather a unified *proxy-tactile perception* of the robot peripersonal space.

This large-area sensing architecture provides comprehensive spatial awareness of the environment, enabling robots to detect objects and humans before contact occurs, interpret physical interaction when contact happens, and adapt their behavior accordingly [14], [15].

The SESTOSENSO framework presented in this paper builds upon this concept. At a conceptual level, it is organized into three main layers, as illustrated in Fig. 1. The first layer consists of the distributed proxy-tactile sensing infrastructure provided by ProxySKIN, which integrates tactile and proximity sensing elements over the robot body. This layer also includes the embedded processing and communication modules responsible for sensor data acquisition, low-level preprocessing, and transmission. The second layer comprises the perception and control modules that exploit proxy-tactile information to perform environment modeling, interaction interpretation, and reactive control for safe human-robot collaboration. In particular, tactile and proximity measurements acquired by ProxySKIN are locally processed by the embedded units, interpreted by dedicated perception algorithms, and finally exploited by control strategies to regulate robot motion and physical interaction. The third layer is represented by the application modules, which exploit the underlying hardware and software components to implement task-specific functionalities in collaborative scenarios. These applications validate the proposed technologies in representative use cases and demonstrate their integration within realistic robotic workflows.

The following subsections describe the main architectural components of the SESTOSENSO framework, with particular emphasis on the ProxySKIN sensor network and the embedded processing infrastructure enabling distributed proxy-tactile perception.

III. Proxy-Tactile System Architecture

At the lowest architectural level, the system consists of the ProxySKIN sensor network and the embedded processing units associated with each sensing module. The former provides distributed tactile and proximity sensing over the robot body, whereas the latter performs low-level acquisition, preprocessing, and communication with the higher-level perception and control layers.

A. ProxySKIN Sensor Network

The ProxySKIN technology [16], as illustrated in Fig. 2, provides a distributed multimodal sensing layer covering large portions of the robot body. Each sensing module integrates capacitive tactile transducers and miniaturized Time-of-Flight (ToF) proximity sensors within a flexible architecture conformable to curved robot surfaces, such as manipulator links. The tactile sensing layer is based on the CySkin technology, which employs distributed capacitive taxels arranged over flexible printed circuit boards. Each tactile module integrates

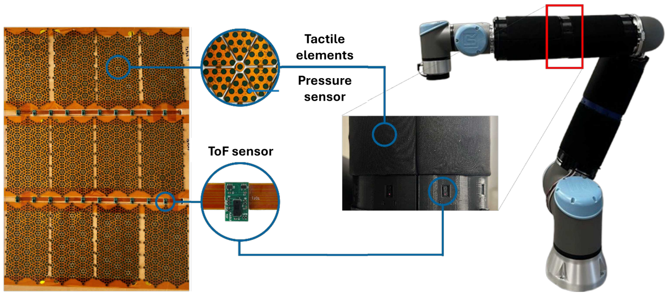

Fig. 2. Integration of the ProxySKIN technology onto a UR10e manipulator. The system combines distributed capacitive tactile elements and embedded miniaturized proximity sensors within flexible sensing modules conformable to curved robot surfaces.

11 capacitive taxels based on 3.5 mm conductive pads, providing an overall tactile spatial resolution of approximately 2.03 pressure sensors/cm^2, with acquisition frequencies up to 20 Hz. The ToF sensing layer relies on VL53L8CX sensors providing 8×8 multizone ranging measurements (64 depth samples per frame), a sensing range up to 4 m, and acquisition frequencies up to 15 Hz in the 64-zone configuration.

The tactile sensors measure the pressure distribution generated during physical interactions, whereas the ToF sensors provide distance measurements for the early detection of nearby obstacles. These two sensing modalities are conceived as complementary technologies, enabling both contact and pre-contact perception to enhance safety in human–robot collaboration scenarios, in line with the principles of Industry 4.0 and the emerging vision of Industry 5.0. The proposed architecture supports the deployment of large-area sensing skins while preserving the robot mechanical compliance and mobility.

Furthermore, previous characterization studies [16] demonstrated that interference effects among adjacent ToF sensors remain negligible and affect distance measurements only at the millimeter scale, which is acceptable for typical human–robot interaction applications. Although the adopted infrared-based ToF technology enables distributed pre-contact perception, sensing performance can degrade in presence of transparent or highly reflective materials such as glass or mirrors, depending on the environmental conditions and target optical properties.

Nevertheless, compared with traditional visual-servoing architectures, distributed sensing technologies offer large-area perception capabilities, thereby reducing the occlusion issues that commonly affect both eye-to-hand and eye-in-hand configurations due to the limited field of view of cameras [17]–[19]. This aspect is particularly important in unstructured and cluttered environments, where even small losses of perceptual information may critically compromise safety in industrial collaboration and cooperation tasks. In addition, the proposed technology provides a higher degree of modularity than camera-based systems. Finally, it also mitigates the privacy concerns associated with camera-based monitoring, an issue that is receiving increasing attention from both industry and society [20], [21].

B. Embedded Sensor Processing

Each sensing module is connected to embedded acquisition units responsible for local data processing and communication with the robot control system. These units perform sensor readout, synchronization, and preliminary filtering of tactile and proximity measurements before transmitting the data to the higher-level perception modules.

The distributed processing architecture reduces the amount of raw data transmitted over the network and improves the scalability of the sensing system. Sensor nodes communicate with the robot controller through a lightweight communication layer, allowing multiple sensing patches to be integrated across the robot body while minimizing cabling complexity.

IV. Proxy-tactile perception framework

The proxy-tactile perception framework exploits distributed tactile and proximity sensing to derive structured representations of the environment and of the interaction dynamics around the robot. By fusing ProxySKIN measurements with proprioceptive information and machine learning-based processing, the framework enables several complementary perception functions, ranging from environment modeling and self-localization to human-robot interaction, tactile gesture recognition, and motion estimation in the surrounding scene.

A. Environmental Perception and Self-Localization

Proxy-tactile sensing data are fused with proprioceptive measurements to construct robot-centric representations of the surrounding environment. The adopted approach represents environmental information through point clouds encoding both geometric and mechanical properties. Each of the 64 distance measurements acquired by a ToF module is individually converted into a three-dimensional point through geometric projection based on the sensor intrinsic parameters and its pose relative to the robot structure. The resulting points are then transformed into the robot base frame, enabling the construction of a spatially coherent representation of the environment. By aggregating the measurements from distributed sensors, the system reconstructs a sparse yet consistent point cloud describing obstacles and surfaces in proximity to the robot.

The proximity-based representation can be enriched with complementary tactile information acquired during physical interaction. In particular, as discussed in [22], tactile sensing enables the estimation of local mechanical properties of contacted surfaces, such as compliance and contact normals, by relating the applied force to the resulting deformation. Tactile measurements can therefore be encoded as attributes associated with the corresponding points in the geometric map, yielding an augmented point cloud that jointly represents geometric and physical properties of contact sources. The tactile attributes estimated at contact locations can also be propagated to neighbouring regions of the object surface, allowing the robot to reason not only about geometry but also about material characteristics.

In this context, the distributed tactile sensing solution enables the robot to detect and manage multiple simultaneous

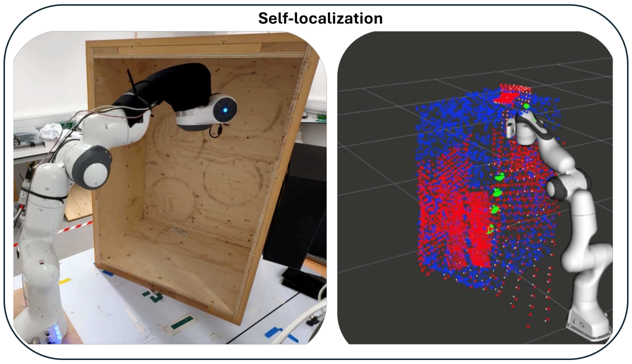

Fig. 3. Estimation of the crate pose through a particle filter driven by measurements collected from miniaturized ToF sensors integrated on the robot manipulator. The environment map, known a priori, is represented with blue dots, while the red point cloud corresponds to the accumulation of point cloud samples acquired at different robot poses.

contacts over its body, unlike traditional force-torque sensing systems. This capability allows the system to reason about complex multi-contact interaction scenarios. By distributing sensing elements over large areas, the robot can maintain low interaction forces while still acquiring meaningful contact information, thereby enabling safe exploration of cluttered environments and physically safe interactions with human operators [23]. Safety can be further enhanced by reconstructing the local geometry of objects in contact with the tactile sensors and by exploiting convolutional neural networks to recognize the contacted objects [24], [25]. In this way, the system can trigger context-dependent robot behaviors and support a form of non-verbal interaction between the operator and the robot.

Beyond environment modeling, ProxySKIN can also support robot self-localization. In Figure 3, by exploiting the spatial distribution of the proximity sensors over the robot body, the system can estimate the relative position of nearby structures and track their motion over time [26]. The resulting robot-centric map provides a reference that allows the robot to infer its pose with respect to surrounding objects, even in scenarios where conventional vision-based localization may be unreliable because of occlusions or limited visibility.

B. Human–Robot Interaction

A key objective of the project is to enable safe and intuitive human-robot interaction. To this end, the developed perception system combines proximity sensing, tactile feedback, and machine learning techniques to detect human presence and recognize gestures in order to trigger specific robot behaviors.

Within this framework, a three-layer Convolutional Neural Network (CNN) is employed to recognize hand contacts and the associated tactile gestures from tactile images generated by the distributed sensing system. The proposed CNN was trained using a dataset composed of more than 1800 tactile contact images collected from 43 different subjects interacting with the robot under multiple contact conditions. The recognized tactile interactions are then exploited to activate subsequent collaboration phases during industrial cooperative tasks, as depicted in Fig. 4.

At the same time, proximity feedback is used to adapt the robot position and behavior so as to reduce the risk of collisions with nearby objects and maintain a predefined safety distance from the human operator during cooperation. This strategy enforces safety as a primary objective, improving operator comfort and perceived trust [15] while preserving task performance.

C. Tactile Gesture Recognition

Tactile sensing enables humans to interact with robots through physical gestures applied to the robot surface. The pressure distribution captured by the tactile sensors can be represented as a sequence of tactile images describing the evolution of contact during the gesture.

However, different gestures may generate similar contact shapes, making them difficult to distinguish using tactile images alone. To address this limitation, the dynamics of the contact can be extracted by computing dense optical flow between consecutive tactile images. The optical flow representation encodes the motion of pressure patterns across the tactile surface and highlights differences between gestures that produce similar contact shapes but different contact dynamics.

The resulting representation combines the tactile image and the optical flow information into a three-channel image encoding contact shape, flow magnitude and flow direction. This augmented representation [27] provides a richer description of the interaction by explicitly capturing both the spatial distribution of the contact and its temporal evolution. In particular, while the tactile image describes the instantaneous pressure pattern applied by the user, the optical flow encodes the displacement of these pressure patterns between consecutive frames, revealing the dynamic characteristics of the gesture. This explicit representation of contact dynamics naturally connects tactile gesture analysis with motion estimation frameworks commonly used in visual perception.

D. Scene Flow Estimation

Proximity sensing from distributed ToF sensors can be used to estimate motion in the robot surroundings. At each time step, the sensors acquire depth measurements that can be converted into a sparse point cloud representing nearby objects in the environment.

Given the distributed placement of the sensors across the robot body, individual measurements are first transformed into a common reference frame and merged, resulting in a global representation of the surrounding scene. However, miniaturized ToF sensors provide low spatial resolution and relatively noisy measurements, producing point clouds that are significantly sparser and less reliable than those obtained with conventional depth cameras or LiDAR systems.

In this context, to estimate the motion of objects in the environment, the scene flow problem is formulated as a geometric registration task between point clouds acquired at different time instants [28]. In particular, the method compares

Fig. 4. A physical human–robot interaction scenario with tactile gesture recognition. Both robot links are equipped with ProxySKIN technology.

the current point cloud with a previously acquired one to estimate the rigid transformation that best aligns the two observations. To facilitate correspondence estimation under sparse measurements, the merged point clouds are first partitioned into spatial clusters using the HDBSCAN algorithm, allowing geometrically coherent groups of points to be processed independently.

The HDBSCAN parameters were experimentally tuned according to the sparse and noisy characteristics of the acquired ToF data. In particular, the minimum cluster size was set to $N_{min} = 60$ to reject isolated noisy measurements while preserving clusters associated with relevant moving objects. For each cluster, the rigid transformation between consecutive point clouds is estimated through the ICP algorithm. Additional filtering stages are introduced to improve robustness, including an ICP fitness threshold $\gamma = 0.7$ to reject unreliable correspondences and a displacement threshold $\delta = 0.04$ m to distinguish actual object motion from noise-induced variations. The temporal separation between the compared point clouds was set to $n = 8$ frames as a trade-off between motion observability and estimation latency [28].

Once a reliable transformation is obtained, point-wise displacements are converted into velocity vectors by considering the temporal interval between the two observations, providing an estimate of the local scene flow. By iteratively applying this procedure over successive acquisitions, the system enables continuous estimation of both the direction and magnitude of motion of objects within the robot's surroundings, even in the presence of sparse and noisy measurements.

V. Proxy-Tactile Reactive Control

Proxy-tactile sensing enables control strategies that exploit both proximity and tactile information to regulate robot behaviour. The proposed proxy-tactile framework enables multiple layers of safety by exploiting both proximity and tactile sensing distributed across the robot body. In Fig. 5 proximity sensing provides early detection of nearby obstacles, allowing the robot to react before physical contact occurs, while tactile sensing allows the system to safely manage interactions when contact cannot be avoided.

The control architecture is implemented using a task-priority framework in which different objectives are assigned different priorities. The highest-priority task enforces the UR10e joint limits and guarantees safe robot configurations during operation. A secondary task exploits proximity measurements obtained from distributed Time-of-Flight sensors to generate reactive avoidance behaviors that increase the distance between the robot and nearby obstacles. Finally, a lower-priority task controls the end-effector motion required to execute the desired manipulation objective.

This hierarchical formulation allows the robot to continue executing the task while dynamically reacting to environmental changes. When obstacles are detected in advance, the robot modifies its motion to avoid contact.

Compared with classical safety approaches such as Speed and Separation Monitoring (SSM), which typically require the robot to stop when humans or obstacles are detected in close proximity, the proposed proxy-tactile strategy enables safe physical interaction by actively controlling contact forces rather than preventing contact altogether.

A. Whole-Arm Avoidance

Distributed proximity sensing further enables whole-arm avoidance behaviours that extend beyond the sensorized regions of the robot. Each ToF sensor produces a set of distance measurements that can be converted into a sparse point cloud representing nearby obstacles in the robot workspace.

These measurements are first processed through a filtering pipeline that removes points belonging to the robot body itself, which may appear in the sensors' field of view depending on the robot configuration. This filtering step relies on the robot geometric model and ray–mesh intersection tests to distinguish measurements originating from the robot structure from those corresponding to external obstacles. The resulting environment point cloud therefore contains only relevant observations of objects in proximity to the manipulator. The resulting point cloud and the robot links, which are also represented as point clouds, are used as inputs to the minimum-distance point pair computation [14]. The identified pair is then used to generate a reactive avoidance motion that pushes the robot away from the detected obstacle while maintaining a safety margin.

The avoidance behaviour is integrated within the task-priority control framework [29] so that obstacle avoidance acts as a secondary objective while still allowing the robot to perform its primary manipulation task. Within this framework, geometric information about the robot body is combined with the proximity measurements to generate avoidance actions on arbitrary points of the robot model, not only at the sensor mounting locations. This capability enables the robot to react to obstacles even when the closest link is not directly equipped with sensors, effectively extending collision avoidance behaviour to nonsensorized parts of the manipulator and enabling whole-arm reactive motion with only partial sensor coverage.

Together with the tactile-based force regulation described in the previous section, this approach provides complementary safety mechanisms: proximity sensing minimizes the likelihood of contact, while proxy-tactile control ensures that

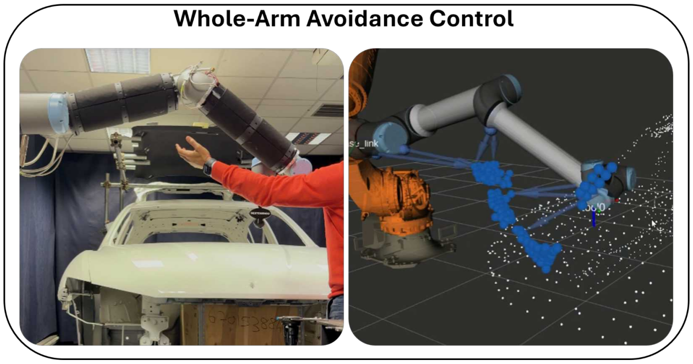

Fig. 5. Real and simulated scenarios of whole-arm avoidance control. The blue points represent the point cloud of the operator's arm approaching the robot, while the blue arrows indicate the minimum-distance point tasks activated on the links.

unavoidable physical interactions are handled safely without requiring the robot to stop.

VI. Results in the Automotive Collaborative Assembly Use Case

To validate the effectiveness of ProxySKIN and the proposed proxy-tactile perception framework, the SESTOSENSO project developed an automotive collaborative assembly demonstrator for the installation of the Maserati Levante headliner in collaboration with Centro Ricerche Fiat (CRF). The demonstrator employed a robotic platform composed of a sensorized UR10e collaborative robot, equipped with 60 proximity sensors and more than 4500 distributed tactile sensors, mounted on a KUKA KR150 R3100-2 heavy-duty industrial manipulator. Within this configuration, the UR10e performed headliner handling, positioning, fixturing assistance, and cooperative assembly tasks, while the KUKA robot extended the workspace and supported component insertion inside the vehicle chassis.

The assembly operation was organized into four phases, described in Fig. 6.

Phase 1 – Headliner pickup and insertion:
The KUKA robot performed a predefined motion toward the support structure holding the headliner, while the UR10e remained in a fixed fully extended posture and carried a pneumatic vacuum gripper. After grasping the component, the robotic system lifted it and introduced it into the chassis through the windshield frame.

Phase 2 – Cooperative roof position adjustment and fixturing:
After insertion, the human operator checked the alignment of the headliner with respect to the chassis and refined its position by physically guiding the UR10e body. The robot admittance controller was activated through the classification of human-hand contacts from images generated by the distributed tactile sensors. During this stage, the robot maintained the component in the desired pose while the operator carried out the initial fixturing steps. The switch to the next phase was commanded through a gesture identified by the distributed proximity sensing system.

Phase 3 – Robot tool change:
The two manipulators followed predefined exit trajectories to leave the vehicle interior and replace the vacuum gripper with a mock screwdriver. Meanwhile, the operator prepared the remaining assembly parts by collecting bolts and internal handles from a nearby bench.

Phase 4 – Cooperative roof screwdriving:
In the last phase, the human and the robot operated simultaneously inside the vehicle to complete the assembly. The operator worked on the left side of the roof, fastening the handles, while the robotic system reproduced the screwdriving procedure on the right side. Safe execution was achieved through a Task Priority Control scheme implemented on the UR10e. In this framework, the highest-priority task enforced safety by adapting robot speed according to the distance from the human estimated through the distributed proximity sensors, while a lower-priority task controlled the end-effector motion required to accomplish the screwdriving operation.

Validation tests were conducted in a laboratory environment designed to reproduce representative industrial conditions. During the experiments, the robotic platform safely operated in the vicinity of both human operators and surrounding structural elements. The combined use of tactile and proximity sensing enabled the detection of nearby obstacles and contact events during task execution. Accidental impacts and damage occurrence remained below the project threshold of 5%, with an experimentally measured damage rate of approximately 3%.

The experimental campaign included insertion and manipulation tasks involving rigid, compliant, and deformable components. A summary of the quantitative validation results is reported in Table I. A total of 60 validation trials were performed across different operating conditions and configurations, including manual execution performed only by the human operator, operation with the cobot alone, and collaborative human–cobot execution. Repeatable performance was observed across consecutive trials, and no major faults occurred during the validated experimental sequences. The system successfully handled all target object classes, while the overall task completion rate reached 100%, exceeding the project target of 90%. Reliability analysis showed only 2 errors over the 60 executed trials, corresponding to an error/damage rate of approximately 3%, remaining below the target threshold of 5%.

The measured cycle times, compared against the revised manual baseline procedure, achieved an approximately 6% reduction, slightly surpassing the expected 5% target. These results indicate that robotic assistance can be introduced without penalizing process throughput while maintaining safe operation during collaboration.

In terms of human interaction, the use of the cobot required limited adaptations in the operator's movements. The repeatability and predictability of the robotic behavior contributed to stable interaction patterns during collaboration. Furthermore, TAM, NASA-TLX, SUS, and TRUST assessments indicated

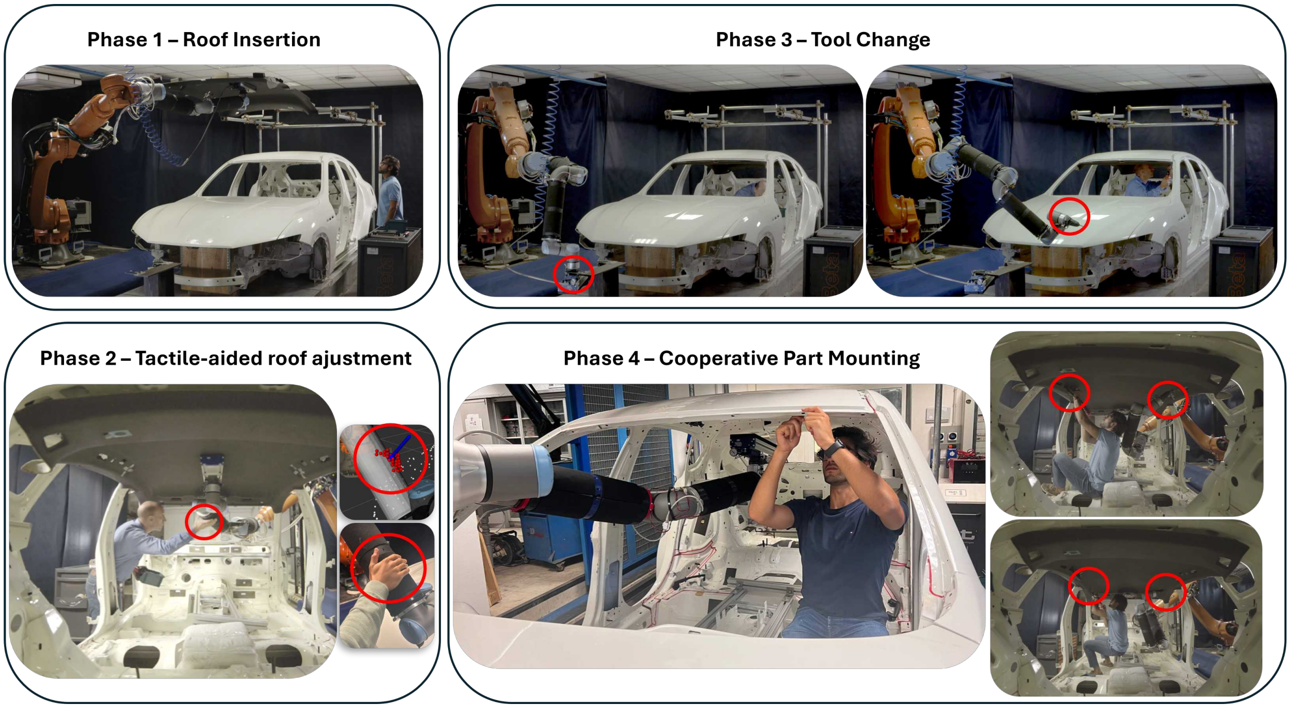

Fig. 6. The complete automotive demonstrator representing the four phases, highlighting the tool change, the phases of physical human–robot interaction and the collaborative phases within the vehicle chassis.

positive user acceptance and no evidence of increased mental workload [15].

The ToF sensing layer provided distributed perception of the space surrounding the robot body, whereas the tactile sensing system enabled reliable surface contact detection during physical human–robot interaction.

TABLE I
SUMMARY OF THE RESULTS OBTAINED IN AUTOMOTIVE SCENARIO

Metric	Experimental Result
Validation trials	60
Task completion rate	100%
Damage / error rate	$\sim 3\%$
Number of recorded errors	2 / 60 trials
Cycle time improvement	$\sim 6\%$ reduction
NASA-TLX, SUS, and TRUST assessments	Positive acceptance

VII. DISCUSSION

The experimental results demonstrated that the proposed proxy-tactile framework can effectively support collaborative automotive assembly operations while maintaining acceptable safety and interaction performance. The integration of distributed tactile and proximity sensing enabled both pre-contact awareness and physical interaction handling, achieving a 100% task completion rate, a damage rate of approximately 3%, and an approximately 6% cycle-time improvement with respect to the revised manual baseline procedure. These results are consistent with the validation objectives and KPIs defined within the SESTOSENSO project framework.

Nevertheless, the proposed approach still presents several limitations. The miniaturized ToFs provide sparse and relatively noisy measurements compared with conventional depth cameras or LiDAR systems, potentially reducing perception accuracy in cluttered environments and in presence of transparent or highly reflective surfaces such as glass or mirrors. Additional limitations are related to the responsiveness of the perception and control pipeline, particularly during highly dynamic human–robot interaction scenarios. Although the distributed multimodal architecture partially mitigates these issues through tactile feedback integration, further improvements in sensing robustness, spatial resolution, sensor fusion, and predictive control strategies remain important directions for future work.

From the human–robot interaction perspective, the repeatability and predictability of the robotic behavior contributed to stable collaboration patterns and positive user acceptance, as reflected by the TAM, NASA-TLX, SUS, and TRUST evaluations. Overall, the experimental campaign confirmed that the proposed framework successfully achieved the primary research objectives related to safe human–robot collaboration, distributed proximity awareness, and cooperative task execution within the considered automotive assembly scenario.

VIII. CONCLUSIONS

This paper presented the vision and architecture of the SESTOSENSO framework for proxy-tactile perception in collaborative robotics. The proposed system integrates distributed tactile and proximity sensing across the robot body with perception and control modules, enabling a unified representation of the robot's peripersonal space and supporting safe and adaptive human–robot interaction.The approach was validated through an automotive collaborative assembly demonstrator, where a sensorized robotic system assisted a human operator during vehicle roof installation. Experimental results confirmed the feasibility and robustness of the proposed architecture, demonstrating reliable perception of nearby obstacles and contact events, safe human–robot interaction, and consistent task execution across repeated trials. The evaluation also showed positive operator acceptance and indicated that robotic assistance can be introduced without compromising process efficiency.

Future work will focus on the tighter integration of sensing, perception, and control modules within a unified cognitive proxy-tactile robotic framework, together with the extension of the proposed technologies to additional industrial scenarios and application domains.

REFERENCES

[1] Z. Saleem, F. Gustafsson, E. Furey, M. McAfee, and S. Huq, "A review of external sensors for human detection in a human robot collaborative environment,"Journal of Intelligent Manufacturing, vol. 36, pp. 2255–2279, 2025, doi: 10.1007/s10845-024-02341-2.

[2] R. Halme, M. Lanz, J. Kämäräinen, R. Pieters, J. Latokartano, and A. Hietanen, "Review of vision-based safety systems for human-robot collaboration," Procedia CIRP, vol. 72, pp. 111–116, 2018, doi: 10.1016/j.procir.2018.03.043.

[3] A. Mohammed, B. Schmidt, and L. Wang, "Active collision avoidance for human–robot collaboration driven by vision sensors," International Journal of Computer Integrated Manufacturing, vol. 30, no. 9, pp. 970–980, 2017, doi: 10.1080/0951192X.2016.1268269.

[4] L. Wang, B. Schmidt, and A. Y. C. Nee, "Vision-guided active collision avoidance for human-robot collaborations," Manufacturing Letters, vol. 1, no. 1, pp. 5–8, 2013, doi: 10.1016/j.mfglet.2013.08.001.

[5] J. De Gea Fernández, D. Mronga, M. Günther, T. Knobloch, M. Wirkus, M. Schröer, M. Trampler, S. Stiene, E. A. Kirchner, V. Bargsten, T. Bänziger, J. Teiwes, T. Krüger, and F. Kirchner, "Multimodal sensor-based whole-body control for human–robot collaboration in industrial settings," Robotics and Autonomous Systems, vol. 94, pp. 102–119, 2017.

[6] P. S. Girão, P. M. Ramos, O. Postolache, and M. Pereira, "Tactile sensors for robotic applications," Measurement, vol. 46, no. 3, pp. 1257–1271, 2013, doi: 10.1016/j.measurement.2012.11.015.

[7] J. J. O'Neill, J. Lu, R. Dockter, and T. M. Kowalewski, "Practical, stretchable smart skin sensors for contact-aware robots in safe and collaborative interactions," in Proc. IEEE Int. Conf. Robot. Autom. (ICRA), 2015, pp. 624–629, doi: 10.1109/ICRA.2015.7139244.

[8] C. Vogel, M. Fritzsche, and N. Elkmann, "Safe human-robot cooperation with high-payload robots in industrial applications," in Proc. 11th ACM/IEEE Int. Conf. Human-Robot Interact. (HRI), 2016, pp. 529–530, doi: 10.1109/HRI.2016.7451840.

[9] A. Hoffmann, A. Poeppel, A. Schierl, and W. Reif, "Environment-aware proximity detection with capacitive sensors for human-robot-interaction," in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS), 2016, pp. 145–150, doi: 10.1109/IROS.2016.7759047.

[10] Y. Ding, F. Wilhelm, L. Faulhammer, and U. Thomas, "With proximity servoing towards safe human-robot-interaction," in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS), 2019, pp. 4907–4912, doi: 10.1109/IROS40897.2019.8968438.

[11] S. Tsuji and T. Kohama, "Proximity and contact sensor for human cooperative robot by combining time-of-flight and self-capacitance sensors," IEEE Sensors Journal, vol. 20, no. 10, pp. 5519–5526, 2020.

[12] X. Ding, J. Guo, Z. Ren, and P. Deng, "State-of-the-art in perception technologies for collaborative robots," IEEE Sensors Journal, vol. 22, no. 18, pp. 17635–17645, 2022, doi: 10.1109/JSEN.2021.3064588.

[13] SESTOSENSO Project, "SESTOSENSO Project," 2022. [Online]. Available: https://sestosenso.eu/. Accessed: Mar. 6, 2026.

[14] S. Borelli, F. Giovinazzo, A. Albini, F. Grella and G. Cannata, "Generating Whole-Arm Avoidance Motion Through Localized Proximity Sensing," in *IEEE/ASME Transactions on Mechatronics*, vol. 30, no. 5, pp. 3988-3999, Oct. 2025.

[15] S. Borelli, S. Morandini, F. Giovinazzo, F. Grella, F. Fraboni and G. Cannata, "Workspace Sharing with Proximity-Aware Robots: a Pilot on User Perspective," in *2025 34th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, Eindhoven, Netherlands, 2025, pp. 1327-1333.

[16] Giovinazzo, F., Grella, F., Sartore, M., Adami, M., Galletti, R., Cannata, G.: "From cyskin to proxyskin: Design, implementation and testing of a multi-modal robotic skin for human-robot interaction". In: Sensors, vol. 24, no. 4 (2024).

[17] K. He et al., "Visibility Maximization Controller for Robotic Manipulation," in *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 8479-8486, July 2022.

[18] L. Zhu, M. Menon, M. Santillo and G. Linkowski, "Occlusion Handling for Industrial Robots," 2020 *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Las Vegas, NV, USA, 2020, pp. 10663-10668.

[19] A. Casalino, S. Guzman, A. Maria Zanchettin and P. Rocco, "Human Pose Estimation in Presence of Occlusion Using Depth Camera Sensors, in Human-Robot Coexistence Scenarios," 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 2018, pp. 1-7.

[20] T. Winkler and B. Rinner, "Security and privacy protection in visual sensor networks: A survey," ACM Computing Surveys, vol. 47, no. 1, Art. no. 2, pp. 1–42, May 2014.

[21] X. Huang, S. Pan, and M. Bennewitz, "Privacy risks of robot vision: A user study on image modalities and resolution," arXiv preprint, arXiv:2505.07766, 2025.

[22] W. Wasko, A.Albini, P.Maiolino, F. Mastrogiovanni, G. Cannata, "Contact modelling and tactile data processing for robot skin", arXiv preprint, arXiv:1809.08211, 2018.

[23] F. Grella, A. Albini, and G. Cannata, "Voluntary interaction detection for safe human-robot collaboration," in *Proc. 6th IEEE Int. Conf. Robotic Comput.*, 2022, pp. 353–359.

[24] A. Albini, S. Denei, and G. Cannata, "Human hand recognition from robotic skin measurements in human-robot physical interactions," in *Proc. 2017 IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Vancouver, BC, 2017, pp. 4348–4353.

[25] A. Albini, G. Cannata, "Pressure Distribution Classification and Segmentation of Human Hands in Contact with the Robot Body", in *The International Journal of Robotics Research (2020)*.

[26] G. Caroleo, A. Albini, D. De Martini, T. D. Barfoot and P. Maiolino, "Tiny LiDARs for Manipulator Self-Awareness: Sensor Characterization and Initial Localization Experiments," 2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hangzhou, China, 2025, pp. 12940-12946.

[27] S. Zhong, A. Albini, G. Caroleo, G. Cannata, and P. Maiolino, "Improving tactile gesture recognition with optical flow," in *Proc. IEEE Int. Conf. on Robot and Human Interactive Communication (RO-MAN)*, Eindhoven, Netherlands, 2025.

[28] J. Sander, G. Caroleo, A. Albini, and P. Maiolino, "Estimating scene flow in robot surroundings with distributed miniaturised time-of-flight sensors," in *2025 34th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, Eindhoven, Netherlands, 2025, pp. 493-499.

[29] E. Simetti, G. Casalino, F. Wanderlingh, M. Aicardi, "Task priority control of underwater intervention systems: Theory and applications," in *Ocean Engineering*, Volume 164, 2018, Pages 40-54.