

# VISTAF RoboSkin: Vision-Integrated Multimodal Sensing of Shape, Temperature, and Force from a Single RGB Image

Rim El Qabli, Supervisor: Benhui Dai, Professor: Josie Hughes  
 Robotics Semester Project, CREATE Lab, EPFL

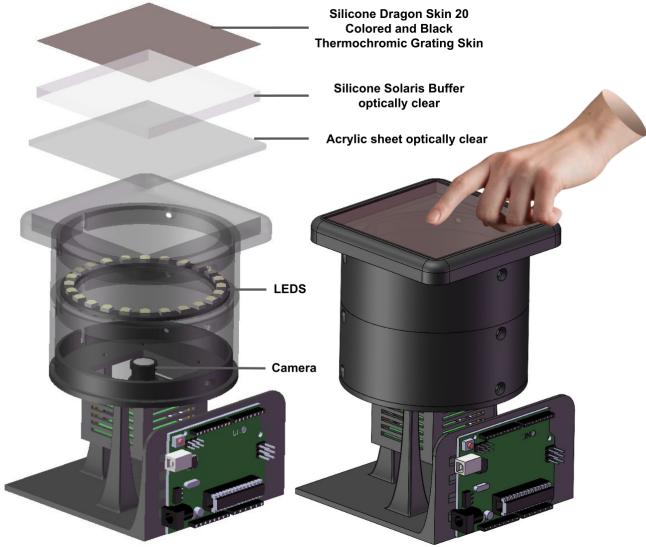


Figure 1. System overview. Vision-based multimodal skin: camera + LED ring observe a 0.6 mm TLC grating through a transparent stack.

**Abstract**—We present VISTAF Roboskin, a compact vision-based soft sensing skin that jointly and passively estimates force, shape and temperature from a single RGB camera view. A thin (0.6 mm) silicone grating with embedded thermochromic pigment (TLC) powders is imaged from below under controlled LED-ring illumination. Deformation is recovered by Fourier Transform Profilometry (FTP) from phase differences between a reference and deformed frame, then mapped to metric depth and force via calibration. Temperature is inferred from TLC appearance using CIELAB-based regression. A two-stripe design (high-sensitivity colored stripes and broad-range black stripes) preserves FTP contrast while expanding thermal coverage, enabling dense depth/force and temperature maps without electronics in the skin. On calibration datasets, phase→height achieves RMSE  $\approx 2.17 \times 10^{-3}$  mm, volume→force achieves RMSE  $\approx 6.96$  N, and mean-per-setpoint temperature prediction achieves RMSE  $\approx 0.44$  °C (colored) and RMSE  $\approx 1.93$  °C (black).

## I. INTRODUCTION

Soft, skin-like sensing has long been central to safe physical interaction in robotics, wearables, and healthcare by providing high-resolution, spatially distributed contact information. Many tactile/thermal skins rely on resistive, capacitive, piezoelectric, or barometric arrays, which often require multilayer stacks, dense routing, and calibration that complicate scaling to palm-sized areas and can reduce compliance [7], [6], [8]. Multimodal electronic skins exist, but cross-sensitivity, drift, and assembly complexity remain practical barriers for robust integration.

Vision-based tactile sensing offers an alternative: a camera observes a deformable medium under controlled illumination, enabling dense deformation maps with minimal embedded electronics [10], [11]. Recent work shows the value of palm-scale vision-based sensing for manipulation [4]. Extending these sensors to *temperature* is less mature. We use thermochromic liquid-crystal (TLC) pigment powders: microencapsulated materials whose visible appearance varies with temperature and, in the formulations used here, progressively fades

to transparent/near-colorless beyond a nominal “clearing” activation temperature, this rather than switching to a fixed new color. In the remainder, we refer to these powders as “TLC pigments” for brevity. Prior vision-based skins have combined deformation and thermal appearance using thermochromic layers [2], [3], but multimodal designs face cross-talk (shading and translucency changes) and fabrication challenges when thin thermochromic coatings must coexist with a stable visual carrier for shape recovery.

This project addresses these gaps with a single TLC-embedded grating skin that jointly supports FTP deformation recovery and temperature estimation. The grating required by FTP is reused as the thermal substrate by mixing TLC powders directly into the silicone. A camera and LED ring view the grating from below: FTP phase differences yield depth/shape and force, while CIELAB regression yields temperature. Two stripe populations, colored stripes tuned for high sensitivity in the touch range and black stripes for broader thermal range coverage, preserve grayscale contrast for FTP [1].

**Contributions.** We contribute: (1) a unified multimodal grating skin that passively co-encodes deformation and temperature within a single continuous soft surface; (2) a reproducible 0.6 mm fabrication method based on direct TLC mixing and two-step casting of colored/black stripe populations; (3) an interpretable pipeline combining FTP (phase→height→force) and CIELAB temperature regression with heating/cooling evaluation; (4) an integrated inference workflow that outputs dense maps and per-session artifacts for traceability; and (5) quantitative calibration performance demonstrating RMSE  $\approx 6.96$  N force estimation and mean-per-setpoint temperature prediction down to RMSE  $\approx 0.44$  °C (colored stripes).

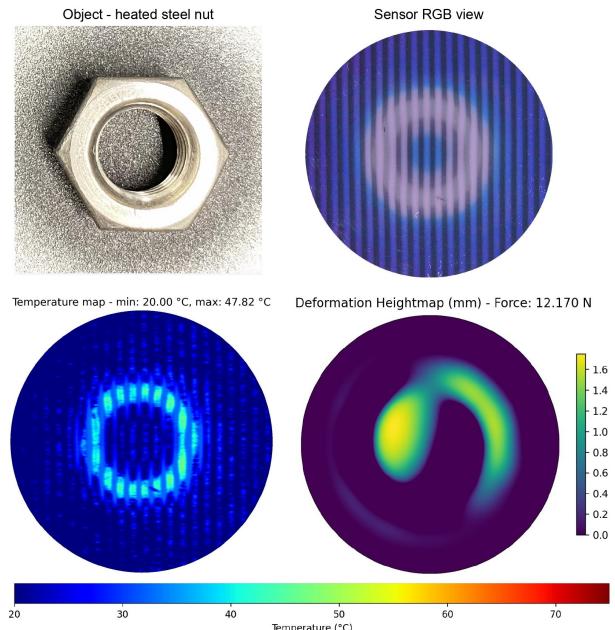


Figure 2. VISTAF Multimodal Sensor demo (cf. VIII-D). A heated steel nut is pressed into the grating skin. From the camera view (reference + deformed), the pipeline reconstructs deformation (heightmap/estimated shape + estimated applied force) and temperature (colormap with min/max temperatures).

## II. SYSTEM OVERVIEW

Fig. 1 summarizes the sensor stack and computation pipeline. A fixed-focus USB camera views the underside of the grating skin through a transparent acrylic window and an optically clear silicone buffer. Illumination is provided by a ring of LEDs integrated into the housing. The pipeline takes (i) a reference image with no contact and (ii) a deformed image under contact and/or heating. From the pair, the system outputs a depth map (FTP) and a temperature map (TLC regression). At each session, a JSON summary stores the results obtained, as well as the model equations and performance metrics for traceability.

A representative multimodal example (heated object + contact) and the resulting force/deformation and temperature reconstructions are shown in Fig. 2.

## III. DESIGN AND FABRICATION

### A. Grating geometry and layer stack

The final skin uses a 2 mm grating period to balance FTP phase sensitivity and manufacturing repeatability. While smaller periods increase carrier frequency and can improve spatial precision, sub-millimeter gratings substantially increase fabrication cost and production time, conflicting with the project goal of a robust, repeatable multimodal prototype.

The final grating thickness is 0.6 mm total:

- **Colored TLC stripes:** a full 0.6 mm thickness of Dragon Skin 20 mixed with a 3-pigment TLC blend (yellow, red, blue). The nominal clearing/activation temperatures are 25 °C, 28 °C, and 31 °C, respectively, yielding high sensitivity in the touch-relevant range and a colorful gradient under heating.
- **Black TLC stripes:** a superposed structure: on top of the 0.2 mm-grooved colored TLC base (continuous to the colored stripes), lie the 0.4 mm black TLC layer. This preserves the global thickness at 0.6 mm while producing visually dark gratings in contrast with the colored stripes and enabling a broader thermal response (8 black TLC pigments with activation set 10 °C, 18 °C, 25 °C, 28 °C, 31 °C, 38 °C, 45 °C, 70 °C).

This two-population design is central: it maximizes temperature information content while preserving grayscale contrast required by FTP. In the TLC formulations used here, the response is characterized by a progressive loss of visible color (increasing transparency) beyond each pigment's nominal activation temperature, which alters both the color and resulting grating contrast under fixed illumination.

### B. Materials and key components

**Elastomers and coatings.** Dragon Skin 20 (Smooth-On) was selected over softer alternatives (e.g., Ecoflex 00-30) due to improved robustness for thin casting and better recovery after repeated indentation, reducing grating drift that degrades FTP reference subtraction. The optically clear buffer (allowing deformation) was fabricated in Smooth-On Solaris (for clarity through the camera). Early prototyping used NOVOCOCS (Smooth-On) to dilute silicone for airbrushing TLC layers, although the final sensor avoided airbrushing for the TLC skin itself but used Psycho Paint (Smooth-On) diluted with NOVOCOCS to create an optically clear adhesive interface and bond the grating skin to the Solaris buffer with minimal air gaps.

**Camera.** The system uses a fixed-focus industrial USB camera (ELP IMX586, up to 8000×6000) to avoid autofocus drift and maintain consistent intrinsics across datasets [5].

### C. Pigment mixing ratios (final TLC skin)

The final TLC grating skin was produced by direct mixing of TLC powders into Dragon Skin 20. The ratios used are referenced in Appendix A.

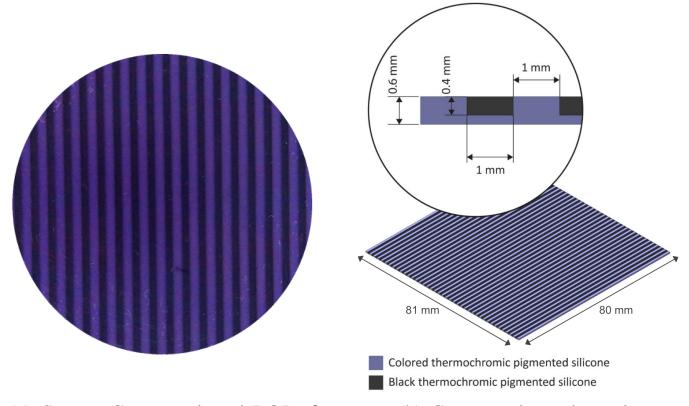
Table I  
MICHELSON CONTRAST SCORES USED TO SELECT THE COLORED TLC COMBINATION.

| Colored Stripe TLC combination | Michelson contrast $C_M$ |
|--------------------------------|--------------------------|
| White                          | 0.93801                  |
| Yellow + Red + Blue            | 0.75183                  |
| Blue + Red                     | 0.48452                  |
| Blue                           | 0.53484                  |

### D. Contrast-driven choice of colored TLC set (brief)

FTP requires a strong grating contrast in grayscale. Because TLC pigments change appearance with temperature, we selected a colored set that maintains high contrast against the black stripes across the working range. Table I lists the reported Michelson contrast scores that guided selection (higher is better).

**Design implication.** The yellow-red-blue combination provides strong chromatic variation between 25 °C to 31 °C while maintaining adequate grayscale separability from dark stripes for FTP. Above the activation range, colored stripes become increasingly translucent, further improving grayscale contrast relative to the black gratings, even as the black gratings also change and brighten with temperature.



(a) Sensor Camera view / ROI of final skin

(b) Cross-section schematic

Figure 3. Final TLC grating skin.

### E. Fabrication iterations and final casting workflow

**Iteration 1: black/white silicone gratings.** We first validated grating fabrication using conventional pigments. A 3D-printed mold formed striped base, while the second material was then squeegeed into grooves to control the thickness. Multiple thicknesses and periods were tested, while 2 mm period was retained for maximized repeatability, precision and affordability.

**Iteration 2: airbrushed TLC coatings (abandoned).** Following prior vision-based thermochromic skins [2], [3], we attempted to airbrush a thin TLC-silicone layer over a black/white grating to retain stripe visibility while adding thermal color. In practice, curing within the airbrush, nonuniform deposition, and masking difficulty for two stripe populations made the process unreliable.

**Final method: direct TLC mixing.** The final skin uses direct mixing of TLC powders into Dragon Skin 20 and two-step casting:

- 1) Cast **colored TLC silicone** into the grating mold to form the base geometry and colored stripes.
- 2) After curing, fill grooves with **black TLC silicone** using a 0.4 mm thickness guide and scraping to leave material only in the grooves, yielding black stripes on top of a 0.2 mm colored base.
- 3) Bond the grating skin to the Solaris buffer using diluted Psycho Paint (airbrushed) for an optically clear interface.



Figure 4. Force experiment setup. 3D-printed testing stand applies load via a hemispherical indenter; a scale measures normal force. 0.5 N to 20 N collected with the stand, 25 N to 45 N applied manually due to stand limits.

#### IV. EXPERIMENTAL SETUP AND DATA COLLECTION

##### A. Palm-like sensing module

The mechanical layout follows a compact palm-sensor style: a rigid housing holds the camera and LED ring at a fixed distance from the grating, viewing through an acrylic window. This design is aligned with palm-scale tactile integration motivations [4]. A fixed-focus camera avoids focus drift and ensures repeatable feature extraction for both FTP and color regression [5]. During processing, a region of interest (ROI) is selected to exclude vignetted edges and LED reflections, which otherwise reduce FTP quality and bias color statistics.

##### B. Force calibration setup (height-to-force)

Force calibration images were captured using a scale-based loading procedure:

- **Controlled force range 0.5 N to 20 N:** a 3D-printed force-testing stand with a large threaded rod applied controlled displacement/force through a hemispherical indenter. The normal force was measured by a scale under the sensor. For each force level, 5 images were recorded to estimate mean and variability.
- **Extended range 25 N to 45 N:** due to stand limitations at higher loads, forces were applied manually. This produces visibly higher variance and is treated as a known limitation (Sec. IX-B).

Fig. 7 shows the volume-to-force fit and the boxplots display variance by force level.

##### C. Temperature calibration setup (heating/cooling)

Temperature calibration used homogeneous 8 cm × 8 cm samples with the same material stack/structure as the final skin, ensuring that optical thickness, pigment density, and backing match inference conditions. Temperature was applied using a controllable heat source, here being a 3D printer bed, recording both heating and cooling sequences to assess hysteresis.

**Stabilization pre-test:** We estimated thermal stabilization time from the convergence of the ROI mean CIELAB  $L$  value over time, then used conservative waiting times to ensure stable color before capture: **4 minutes** for 1 °C steps and **7 minutes** for 5 °C steps. Each time, we took 5 images per step to ensure stability to noise variance.

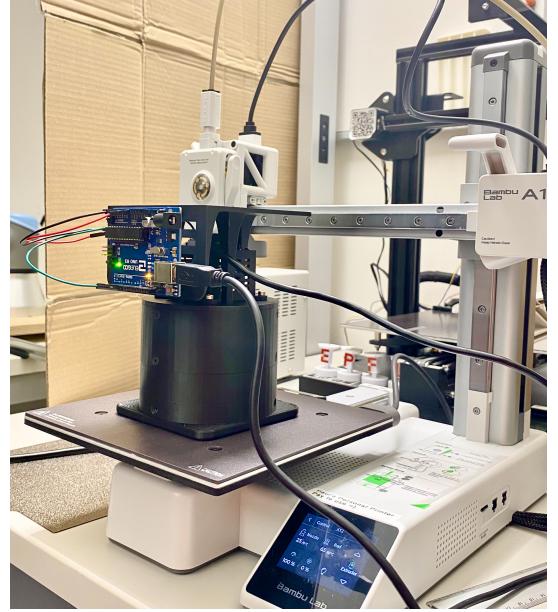


Figure 5. Temperature experiment setup. Homogeneous TLC skin samples (colored and black TLC) are heated/cooled on a 3D printer bed, acting as a controlled temperature source. At each temperature step, the system waits for stabilization, then records 5 images to compute mean and standard deviation.

#### V. METHODS I: FORCE AND SHAPE SENSING VIA FOURIER TRANSFORM PROFILOMETRY

##### A. FTP signal model and processing

Fourier Transform Profilometry (FTP) models the grating as a carrier sinusoid, with the deformation encoded as a spatial phase term [1]:

$$g(x, y) \approx r(x, y) \cos(2\pi f_0 x + \phi(x, y)), \quad (1)$$

where  $f_0$  is the grating spatial frequency and  $\phi(x, y)$  is the deformation-induced phase. We extract  $\phi(x, y)$  by Fourier filtering around the first harmonic and phase unwrapping on the ROI.

##### B. Reference subtraction and unitless height

For robust deformation recovery, we use both a reference image (no contact) and a deformed image (contact). Phase is computed for both frames and subtracted:

$$\Delta\phi(x, y) = \phi_{\text{def}}(x, y) - \phi_{\text{ref}}(x, y),$$

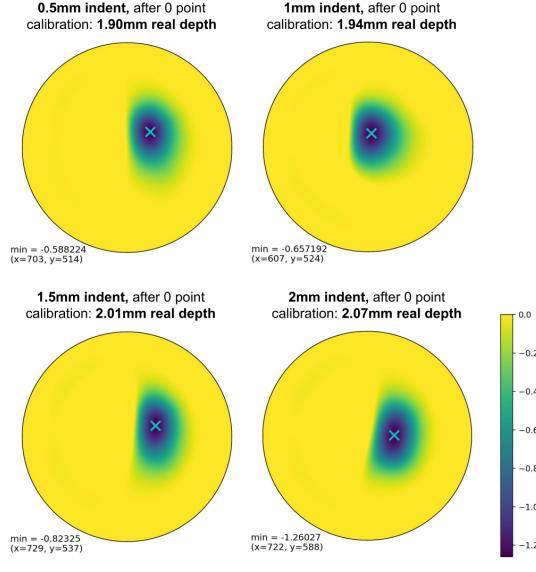
which aims to remove static spatial biases due to illumination nonuniformity, printing imperfections, and assembly geometry. The immediate FTP output is a *unitless height*-like map (proportional to phase change) whose scale depends on the optical geometry and the grating response.

##### C. Calibration A: phase metric → metric depth (mm)

Rather than relying on a full geometric FTP model, we use an empirical mapping from a phase-derived metric to ground-truth indentation depth. Known hemispherical indentation settings (0.5, 1.0, 1.5, 2.0 mm) are used to fit a monotonic saturating relationship.

**Saturating hinge model.** Let  $x$  be a scalar phase-derived metric from the unitless depth map ( $x = -\min(\text{height\_unitless})$ ). We fitted the calibrated depth  $h(x)$  (mm) with:

$$h(x) = H \left( 1 - e^{-k \max(x - x_0, 0)} \right) - H \left( 1 - e^{-k \max(0 - x_0, 0)} \right), \quad (2)$$



(a) Depth maps used for phase→height (mm) calibration

(b) Phase→height (mm) calibration curve

Figure 6. Phase-to-height calibration. Top: reconstructed depth/height maps for 0.5, 1.0, 1.5, 2.0 mm nominal indentations, after enforcing a 0-point reference. Bottom: fitted saturating mapping from the  $x = -\min(\text{height\_unitless})$  phase-derived metric to depth (mm).

where  $H$  is the asymptotic depth,  $k$  controls the rise rate, and the second term enforces  $h(0) = 0$ . Through the enforcing of that 0-point reference, the nominal indentations have been mapped to real perceived indentation depths, mitigating depth distance uncertainties (Fig. 6). The fitted parameter values are provided in Appendix C.

**Calibration accuracy.** On the 4-point depth calibration, phase→height achieves  $R^2 \approx 0.999$  and  $\text{RMSE} \approx 2.17 \times 10^{-3}$  mm.

#### D. Depth map → volume, contact area, and peak depth

After calibration, the pipeline produces a per-pixel depth map  $d(x, y)$  in mm and computes:

- **Max depth**  $d_{\max} = \max d(x, y)$  in the ROI.
- **Contact area**  $A = N_{\text{contact}} \cdot s^2$ , where  $s$  is the scale in mm/px and  $N_{\text{contact}}$  counts pixels exceeding a small threshold.
- **Integrated volume** ( $\text{cm}^3$ ) from the depth field:

$$V_{\text{cm}^3} = \frac{1}{1000} \sum_{(x,y) \in \Omega} [d(x, y) \cdot s^2] \quad \text{for } d(x, y) > \tau, \quad (3)$$

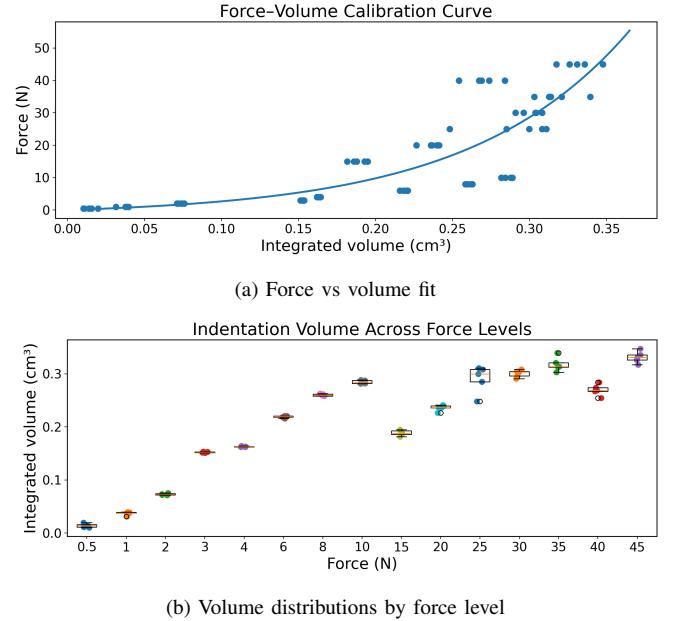
with  $\Omega$  the ROI and  $\tau$  a small depth threshold.

#### E. Calibration B: volume → force

We chose to empirically fit the Force from the scale-based dataset (cf. IV-B) using an exponential growth model:

$$F = \alpha (\exp(\beta V) - 1), \quad (4)$$

where  $F$  is force in N and  $V$  is volume in  $\text{cm}^3$ . The fitted parameter values are provided in Appendix C.



(a) Force vs volume fit  
(b) Volume distributions by force level

Figure 7. Force calibration. Top: fitted exponential relationship between integrated deformation volume and applied force. Bottom: boxplots of volume at each force level (5 images per force), showing increased variance at high forces due to setup limits and manual loading above 20 N.

**Force-model accuracy.** Over the calibrated range (0.5 N to 45 N), volume→force achieves  $R^2 \approx 0.77$  with  $\text{RMSE} \approx 6.96$  N. It is worth noting that variance increases above 20 N due to manual loading (Fig. 7).

## VI. METHODS II: TEMPERATURE SENSING VIA TLC COLOR REGRESSION

### A. Why CIELAB features

TLC pigments encode temperature via appearance changes that are strongly nonlinear and depend on illumination, thickness, and scattering/transparency. We use CIELAB ( $L, a, b$ ) as a more perceptually meaningful representation than raw RGB. For black stripes, we additionally use a grayscale intensity feature (“gray”) to capture appearance changes when chroma is weak or saturates.

### B. Calibration samples and “means” evaluation

At each temperature setpoint, 5 images are recorded after stabilization. For calibration and reporting, we aggregate these 5 frames into a single representative point by averaging the respective metrics, and then evaluating model performance on these **per-setpoint means**. This produces one sample per temperature step ( $n = 14$  for colored,  $n = 36$  for black), aiming to maximize noise reduction.

### C. Model form: polynomial regression

For this part, we trained polynomial regressors using a linear least-squares fit on polynomial feature expansions. Full coefficient values for the color and black models (including separate heating/cooling/global fits) are reported in Appendix C.

**Colored stripes model (global, degree 2).** Using  $\mathbf{z} = [L \ a \ b]^T$ :

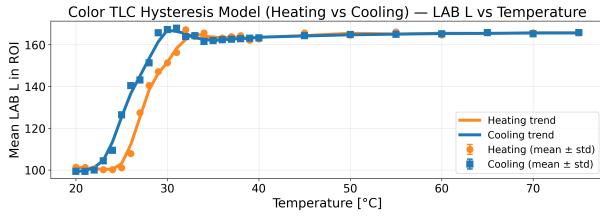
$$T = \mathbf{w}^\top \phi_2(\mathbf{z}), \quad (5)$$

where  $\phi_2 = [1, L, a, b, L^2, La, Lb, a^2, ab, b^2]^T$ .

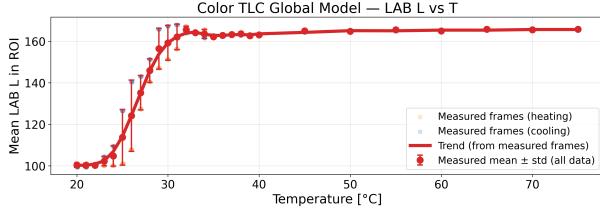
**Black stripes model (global, degree 3).** Using  $\mathbf{u} = [L \ a \ b \ \text{gray}]^T$ :

$$T = \mathbf{v}^\top \phi_3(\mathbf{u}), \quad (6)$$

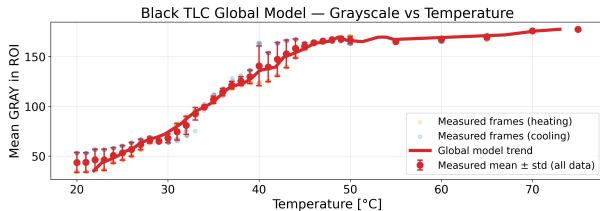
where  $\phi_3$  contains all monomials up to degree 3.



(a) Colored TLC hysteresis Heating vs cooling model ( $L$  vs  $T$ )



(b) Global Merged Colored TLC model ( $L$  vs  $T$  trend)



(c) Global Merged Black TLC model (gray vs  $T$ )

Figure 8. Temperature characterization. Top: Colored-stripe Hysteresis is most visible in the transition region where TLC changes rapidly, but remains negligible, outside that region, heating/cooling trends largely overlap. Middle: The Global merged Colored TLC model exhibits high sensitivity over a narrow range and saturation afterward. Bottom: Global merged Black TLC model provides broader-range encoding with stronger saturation and thus lower precision.

#### D. Trends and global model curves

Fig. 8 provides the key characterization plots:

- **Hysteresis plot (colored):** heating vs cooling  $L$  vs  $T$  with trends and mean $\pm$ std. (NB: Black Hysteresis plot also followed the same principle, and showcased negligible hysteresis as well.)
- **Global  $L$  vs  $T$  (colored):** steep transition over the touch-relevant region (high sensitivity), then saturation at higher  $T$  as TLC becomes translucent.
- **Global gray vs  $T$  (black):** broader-range encoding with stronger saturation regions, explaining reduced precision compared to color model.

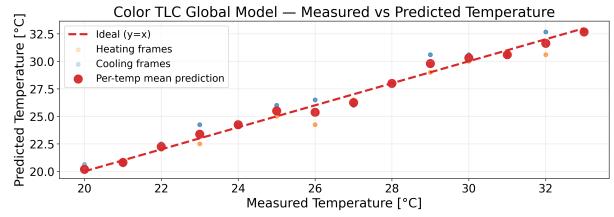
#### E. True vs predicted temperature

Fig. 9 shows parity plots for the global regressors, confirming the trends in Fig. 8: the colored model stays close to the ideal  $y = x$  line over its operating range, while the black model exhibits larger scatter and mild bias in the mid/high range where the TLC response saturates. Small markers correspond to individual heating/cooling frames and large markers to per-setpoint means (5 frames averaged), which are the points used to compute the reported RMSE.

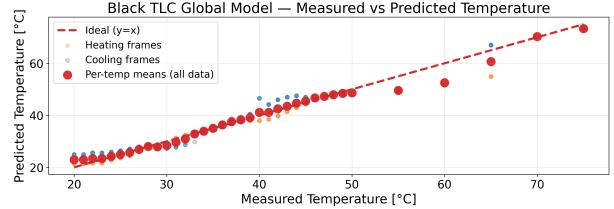
## VII. MULTIMODAL PIPELINE INTEGRATION

### A. Single image-pair inference

The integrated pipeline consumes a **reference** RGB image (no contact) and a **deformed** RGB image (contact and/or heating), and produces (i) a depth map/shape estimate and force estimates from FTP and (ii) a per-pixel temperature map from TLC regression, along with summary statistics. Outputs are saved to a session directory, and a



(a) Global model parity (colored or combined, as exported)



(b) Global model parity (black)

Figure 9. Measured vs predicted temperature. Top: global model parity (colored or combined). Bottom: global model parity (black). Large markers correspond to per-setpoint means (5 frames averaged per temperature step).

JSON summary stores calibration equations, coefficients, and metrics for reproducibility (Appendix C).

### B. Cross-talk management

Cross-talk arises because deformation modifies appearance and temperature modifies contrast. We try to mitigate this primarily through **design** rather than black-box learning:

- 1) Maintain strong grayscale contrast for FTP by choosing pigment mixtures that remain separable across temperature (Table I).
- 2) Use ROI masking and depth thresholding to suppress spurious phase responses at the edges and in reflective regions.
- 3) Train temperature models on samples matched in structure/thickness to inference conditions, reducing systematic bias.

## VIII. RESULTS

### A. Force module results

**Phase-to-height.** The phase-to-height calibration produces a stable monotonic mapping with saturation at larger phase values (Fig. 6). Quantitatively, the fitted mapping achieves  $\text{RMSE} \approx 2.17 \times 10^{-3}$  mm on the depth calibration set.

**Force from volume.** The force-volume relationship is nonlinear (Fig. 7). The boxplots show low variance across the controlled loading region (0.5 N to 20 N) and significantly higher variance in the manual loading region (25 N to 45 N). Overall, the volume $\rightarrow$ force model achieves  $\text{RMSE} \approx 6.96$  N.

### B. Temperature module results

**Colored stripes (high sensitivity, narrow range).** The  $L$  vs  $T$  trend (Fig. 8, middle) shows a steep increase over the touch-relevant range, providing high sensitivity. A slight heating/cooling hysteresis phenomenon can be noticed (Fig. 8, top), although it remains negligible and easily approximated with a global merged model, which trained and chose as the final model (Appendix C).

**Black stripes (broad range, lower precision).** As for the Colored model, the hysteresis model in the Black stripes is deemed negligible, and only the global model is kept. The gray-vs- $T$  plot (Fig. 8, bottom) shows broader coverage than the precise Colored TLC stripes, with stronger saturation regions. This reduces slightly identifiability in some intervals which explains increased prediction error relative to the Colored model.

Table II  
PERFORMANCE SUMMARY (CALIBRATION DATASETS).

| Module                | Model/target                  | Representative performance                                   |
|-----------------------|-------------------------------|--|
| Phase→height          | Hinge-saturating, Eq. (2)     | $R^2 = 0.9988$ ; RMSE = $2.17 \times 10^{-3}$ mm ( $n = 4$ ) |
| Volume→force          | Exponential growth, Eq. (4)   | $R^2 = 0.7745$ ; RMSE = 6.96 N ( $n = 75$ )                  |
| Temperature (colored) | Global poly (deg. 2), Eq. (5) | $R^2 = 0.9882$ ; RMSE = $0.437^\circ\text{C}$ ( $n = 14$ )   |
| Temperature (black)   | Global poly (deg. 3), Eq. (6) | $R^2 = 0.9798$ ; RMSE = $1.93^\circ\text{C}$ ( $n = 36$ )    |

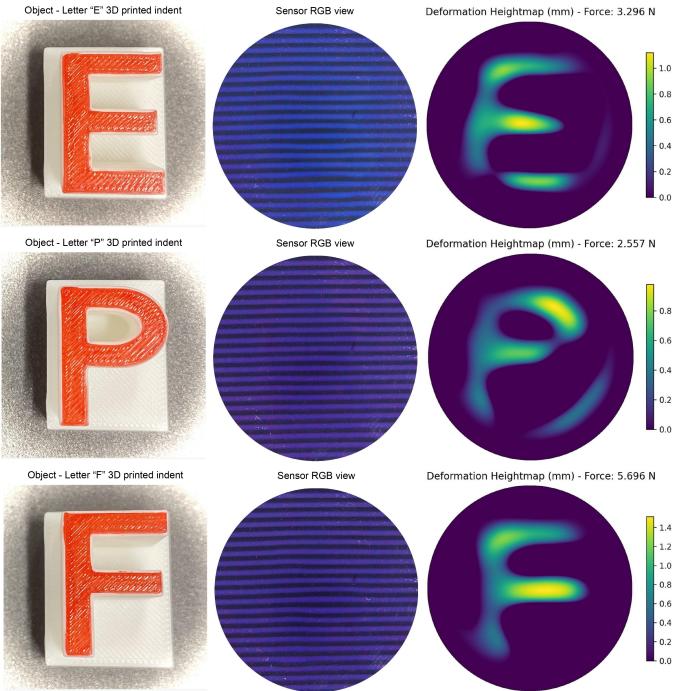


Figure 10. Force/deformation demos with letter imprints (E, P, F). Each panel contains (left-to-right) the object (3D-printed letter indenter), the sensor RGB view, and the reconstructed deformation heightmap with the estimated applied force.

### C. Quantitative performance summary

Table II summarizes calibration performance. As mentioned previously, temperature metrics are reported on per-setpoint means (5 frames averaged per temperature step), reflecting repeatable calibration-curve accuracy.

### D. Demonstrations

We qualitatively validate the multimodal pipeline with three representative demos that span (i) fine shape imprinting and force sensing (Fig. 10), (ii) purely thermal contact with a spatial temperature gradient (Fig. 11), and (iii) multimodal simultaneous thermal + mechanical interaction (Fig. 2, page 1). For each demo, we include an image of the object being pressed on the sensor, together with the corresponding sensor view and the respective reconstructed maps. Limitations displayed by our sensor have been covered in the following section (IX-B).

## IX. DISCUSSION

### A. What worked well

**Interpretability and modularity.** The pipeline separates the vision components cleanly: FTP recovers a dense deformation field from a

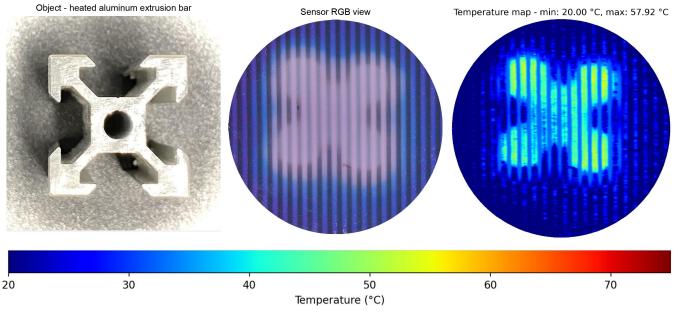


Figure 11. Temperature demo. A heated aluminum bar creates a spatial temperature gradient; the inferred temperature map is recovered from TLC regression (min/max temperatures shown on the colormap).

known grating carrier [1], while temperature is inferred from TLC appearance in perceptual color space. Model coefficients are explicit and reproducible (Appendix C).

**Fabrication repeatability.** Direct mixing of TLC powders into silicone eliminated the largest instability of the airbrushing approach: nonuniform TLC layer thickness and curing inside the airbrush. It also simplified manufacturing and allowed a thin (0.6 mm) skin with fast thermal response.

### B. Limitations

- Force calibration apparatus.** Above 20 N, the test stand could not withstand loads, requiring manual pressure and introducing variance and thus imprecision in the force model (visible in Fig. 7).
- Mechanical stability and registration.** Small shifts between reference and deformed frames can bias phase subtraction. Marker-based registration is a clear next step.
- Optics and ROI.** LED reflections and edge vignetting reduce the usable ROI, setup redesign could increase the usable area.
- Cross-talk under large deformations.** Strong indentation may change local shading and can bias CIELAB features or even alter the precision of the FTP shape and force algorithm (cf. slight shape imprecision, page 1).
- Temperature saturation.** TLC responses saturate outside their active range, as black stripes broaden the range but with reduced precision.

### C. Future work

- 1) Improve mechanical rigidity and add robust registration between frames (for example marker-based registration).
- 2) Improve the FTP-based shape/force pipeline for more robustness.
- 3) Refine grating period while maintaining feasible fabrication.
- 4) Enhance temperature models with deformation-aware stripe segmentation and masking, as well as TLC chroma recovery time study.
- 5) Integrate into robotic hands/palms and leverage dense force/temperature for higher-level perception [4].

## X. CONCLUSION

This work presents a compact vision-based multimodal soft sensing skin that outputs deformation-derived force and shape estimates together with TLC-derived temperature maps from a single RGB camera view. At the heart of our approach is a two-population TLC grating: colored stripes provide high thermal sensitivity, while black stripes extend the usable range, preserving FTP contrast within the same passive skin. On calibration datasets, the system achieves RMSE  $\approx 6.96$  N force estimation and mean-per-setpoint temperature prediction down to RMSE  $\approx 0.44^\circ\text{C}$  (colored) and RMSE  $\approx 1.93^\circ\text{C}$  (black).

## REFERENCES

- [1] M. Takeda and K. Mutoh, "Fourier transform profilometry for the automatic measurement of 3-D object shapes," *Applied Optics*, vol. 22, no. 24, pp. 3977–3982, 1983.
- [2] "A Vision-Based Soft Somatosensory System for Distributed Pressure and Temperature Sensing," *IEEE Xplore*, Art. no. 9001267.
- [3] "HaptiTemp: A Next-Generation Thermosensitive GelSight-Like Visuotactile Sensor," *IEEE Xplore*, Art. no. 9652522.
- [4] "Soft robotic hand with tactile palm-finger coordination," *Nature Communications*, 2025, doi:10.1038/s41467-025-57741-6.
- [5] ELP, "48MP no-distortion USB camera (IMX586) product page," accessed 2025. Available: <https://www.webcamerausb.com/elp-48mp-no-distortion-usb-camera-uhd-8000x6000-cmos-imx586-color-sensor-mini-machine-vision-usb-industrial-camera-uvc-p-482.html>
- [6] J. Qu *et al.*, "Recent progress in advanced tactile sensing technologies for soft grippers," *Advanced Functional Materials*, vol. 33, art. 2306249, 2023, doi:10.1002/adfm.202306249.
- [7] X. Wang, Y. Dong, H. Zhang, R. Yu, C. Pan, and Z. L. Wang, "Recent progress in electronic skin," *Advanced Science*, vol. 2, no. 10, art. 1500169, 2015, doi:10.1002/advs.201500169.
- [8] N. T. Tien *et al.*, "A flexible bimodal sensor array for simultaneous sensing of pressure and temperature," *Advanced Materials*, vol. 26, no. 5, pp. 796–804, 2014, doi:10.1002/adma.201304895.
- [9] N. Wettels, V. Santos, R. Johansson, and G. Loeb, "Biomimetic tactile sensor array," *Advanced Robotics*, vol. 22, no. 8, pp. 829–849, 2008.
- [10] W. Yuan, S. Dong, and E. H. Adelson, "GelSight: high-resolution robot tactile sensors for estimating geometry and force," *Sensors*, vol. 17, no. 12, art. 2762, 2017, doi:10.3390/s17122762.
- [11] M. Lambeta *et al.*, "DIGIT: a novel design for a low-cost compact high-resolution tactile sensor with application to in-hand manipulation," *IEEE Robotics and Automation Letters*, vol. 5, no. 3, pp. 3838–3845, 2020, doi:10.1109/LRA.2020.2977257.

## ACKNOWLEDGEMENTS

We would like to heartfully thank Benhui Dai for his assistance and supervision throughout this project, as well as for his extensive practical support, including CAD assistance and 3D printing/fabrication guidance.

We would also like to thank the CREATE Lab for hosting this project and providing access to all the prototyping and fabrication resources.

## APPENDIX A.

### MATERIAL FORMULATIONS AND MIXING RATIOS

Table III  
FINAL TLC SILICONE MIXING RATIOS USED FOR THE GRATING SKIN.

| Colored stripes (0.6 mm)        | Black stripes (0.4 mm black on top of the 0.2 mm colored base)                        |
|---------------------------------|---|
| Dragon Skin 20: 48 g            | Dragon Skin 20: 32 g  |
| TLC yellow (act. 25 °C): 0.95 g | TLC black pigments: 0.50 g each activation temperature                                |
| TLC red (act. 28 °C): 0.65 g    | Exception: 0.60 g for 25 °C, 28 °C, 31 °C blacks                                      |
| TLC blue (act. 31 °C): 0.60 g   | Activation temperatures mixed: 10 °C, 18 °C, 25 °C, 28 °C, 31 °C, 38 °C, 45 °C, 70 °C |

## APPENDIX B. CODE, DATASET, CAD, AND OUTPUT ARTIFACTS

- **Code repository:** containing the Code, Dataset, Setup CAD and key outputs saved that are showcased in the report's demos (force and temperature outputs, and a JSON summary storing calibration equations, coefficients, and metrics):  
<https://github.com/rimelq/VISTAF-RoboSkin-Vision-Integrated-Multimodal-Sensor/tree/main>.

APPENDIX C.  
COMPLETE MODEL PARAMETERS AND COEFFICIENTS

This appendix lists the fitted parameter values and complete explicit equations used by the pipeline for the final multimodal build.

*A. Phase metric → depth (hinge-saturating)*

Using  $x = -\min(\text{height\_unitless})$ :

$$h(x) = H \left( 1 - e^{-k \max(x-x_0, 0)} \right) - H \left( 1 - e^{-k \max(0-x_0, 0)} \right), \quad (7)$$

with fitted parameters:

$$H = 2.08265, \quad k = 4.20441, \quad x_0 = -1.76784 \times 10^{-9}.$$

(Here the second term enforces  $h(0) = 0$ .)

*B. Volume → force (exponential growth)*

Using  $V$  in cm<sup>3</sup>:

$$F = \alpha (\exp(\beta V) - 1), \quad (8)$$

with fitted parameters:

$$\alpha = 1.61977, \quad \beta = 9.75663.$$

*C. Temperature models (CIELAB regression)*

Below,  $L, a, b$  denote CIELAB channels, and `gray` denotes the grayscale feature used for black stripes.

*1) Colored stripes: heating model (degree 2):*

$$\begin{aligned} T = & 13.848239 + (13.847654) \cdot 1 \\ & + (5.2531972) L + (1.5862391) a \\ & + (0.5783954) b + (-4.4905823) L^2 \\ & + (-4.6004243) La + (2.3772648) Lb \\ & + (-0.81142852) a^2 + (0.75382207) ab \\ & + (0.03512124) b^2. \end{aligned}$$

*2) Colored stripes: cooling model (degree 1):*

$$\begin{aligned} T = & 13.230729 + (13.230712) \cdot 1 \\ & + (4.9260004) L + (1.0703982) a \\ & + (-0.81584244) b. \end{aligned}$$

*3) Colored stripes: global model (degree 2):*

$$\begin{aligned} T = & 13.393386 + (13.393225) \cdot 1 \\ & + (7.948437) L + (4.7591152) a \\ & + (1.8206048) b + (-10.374143) L^2 \\ & + (-15.182861) La + (3.4476021) Lb \\ & + (-3.8808357) a^2 + (4.6297725) ab \\ & + (1.1042786) b^2. \end{aligned}$$

*4) Black stripes: heating model (degree 1):*

$$\begin{aligned} T = & 19.280082 + (19.280062) \cdot 1 \\ & + (27.492858) L + (5.0211744) a \\ & + (1.0514161) b + (-16.596972) \text{gray}. \end{aligned}$$

*5) Black stripes: cooling model (degree 2):*

$$\begin{aligned} T = & 18.078648 + (18.078223) \cdot 1 \\ & + (16.67392) L + (2.6623089) a \\ & + (-0.78947611) b + (-7.082836) \text{gray} \\ & + (-25.217161) L^2 + (-23.559426) La \\ & + (19.890613) Lb + (0.40273462) L \text{gray} \\ & + (0.0041977641) a^2 + (0.52450057) ab \\ & + (25.243333) a \text{gray} + (0.13515792) b^2 \\ & + (-20.48155) b \text{gray} + (26.401713) \text{gray}^2. \end{aligned}$$

*6) Black stripes: global model (degree 3):*

$$\begin{aligned} T = & 17.597937 + (17.59753) \cdot 1 \\ & + (45.091062) L + (1.1637504) a \\ & + (0.55725989) b + (-36.119964) \text{gray} \\ & + (10.036862) L^2 + (11.3199) La \\ & + (0.27752597) Lb + (1.5114454) L \text{gray} \\ & + (0.77557779) a^2 + (-0.25335718) ab \\ & + (-8.6103578) a \text{gray} + (0.25685914) b^2 \\ & + (-0.90476708) b \text{gray} + (-8.7727634) \text{gray}^2 \\ & + (-0.006049401) L^3 + (-0.67782033) L^2 a \\ & + (9.3702651) L^2 b + (-1.3843698) L^2 \text{gray} \\ & + (-10.097261) La^2 + (8.4696894) Lab \\ & + (0.33229855) La \text{gray} + (-2.0422847) Lb^2 \\ & + (-0.74742273) Lb \text{gray} + (-0.25279791) L \text{gray}^2 \\ & + (-0.51770604) a^3 + (0.40586532) a^2 b \\ & + (10.58245) a^2 \text{gray} + (0.14515612) ab^2 \\ & + (-8.6049604) ab \text{gray} + (1.8408661) a \text{gray}^2 \\ & + (0.0050798015) b^3 + (2.3334389) b^2 \text{gray} \\ & + (-10.283551) b \text{gray}^2 + (2.9767837) \text{gray}^3. \end{aligned}$$