Revisiting Lesion Tracking in 3D Total Body Photography

Wei-Lun Huang[a,d], Minghao Xue[a,d], Zhiyou Liu[e], Davood Tashayyod[e], Jun Kang[b], Amir Gandjbakhche[d], Misha Kazhdan[a,2], Mehran Armand[c,a,1,2]

[a]*Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA*
[b]*Department of Dermatology, Johns Hopkins School of Medicine, Baltimore, MD, USA*
[c]*Institute for Integrative and Innovative Research, University of Arkansas, Fayetteville, AR, USA*
[d]*Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, MD, USA*
[e]*Lumo Imaging, Rockville, MD, USA*

Abstract

Melanoma is the most deadly form of skin cancer. Tracking the evolution of nevi and detecting new lesions across the body is essential for the early detection of melanoma. Despite prior work on longitudinal tracking of skin lesions in 3D total body photography, there are still several challenges, including 1) low accuracy for finding correct lesion pairs across scans, 2) sensitivity to noisy lesion detection, and 3) lack of large-scale datasets with numerous annotated lesion pairs. We propose a framework that takes in a pair of 3D textured meshes, matches lesions in the context of total body photography, and identifies unmatchable lesions. We start by computing correspondence maps bringing the source and target meshes to a template mesh. Using these maps to define source/target signals over the template domain, we construct a flow field aligning the mapped signals. The initial correspondence maps are then refined by advecting forward/backward along the flow field. Finally, lesion assignment is performed using the refined correspondence maps. We

---

Email addresses: `whuang44@jhu.edu` (Wei-Lun Huang), `mxue7@jhu.edu` (Minghao Xue), `zoey@lumoscan.com` (Zhiyou Liu), `davood@lumoscan.com` (Davood Tashayyod), `jkang60@jh.edu` (Jun Kang), `gandjbaa@mail.nih.gov` (Amir Gandjbakhche), `misha@cs.jhu.edu` (Misha Kazhdan), `marmand@uark.edu` (Mehran Armand)

[1] Corresponding author

[2] Co-senior authors

propose the first large-scale dataset for skin lesion tracking with 25K lesion pairs across 198 subjects. The proposed method achieves a success rate of 90.1% (at 10 mm criterion) for all pairs of annotated lesions and a matching accuracy of 98.1% for subjects with more than 200 lesions.

Keywords: Total body photography, Skin lesion longitudinal tracking, 3D correspondence

1. Introduction

Melanoma is the most deadly form of skin cancer. Tracking the evolution of existing nevi and detecting new lesions across the body is essential for the early detection of melanoma [1]. Manual evaluation of skin lesions by a dermatologist is considered the standard of care. However, in patients with numerous skin lesions, this task becomes challenging and is prone to human error.

Total body photography (TBP) captures the entire body of a patient using 2D images [2] and/or a 3D mesh [3, 4]. As such, TBP can be effective for monitoring the evolution of lesions [5, 6, 7, 8, 9, 10]. From a systematic review, an individual with a large number (>100) of common nevi is usually included in the target population of TBP [6].

Several prior works have proposed the use of 3D meshes for longitudinal tracking of skin lesions [11, 12, 13, 14]. However, they continue to face significant challenges. First, existing methods produce low-accuracy lesion correspondences, attributed to multiple factors such as the limited positional expression of a lesion, regions with non-isometric deformations, and inconsistent texture across scans. Methods proposed in [12, 13, 14] are constrained by their representation of lesions at the resolution of the mesh vertices, thereby making them heavily dependent on the resolution of the mesh. In extreme cases, adjacent lesions will be mapped to the same vertex, reducing tracking accuracy. In addition, approaches from Zhao *et al.* [12] and Ahmedt-Aristizabal *et al.* [13] rely on a coarse correspondence map between the source and target meshes to identify corresponding lesion pairs. Thus, even with an accurate representation of the lesions themselves, correspondences are necessarily imprecise. Huang *et al.* [14] improve the correspondence accuracy by incorporating both shape and texture information. However, their method is sensitive to inconsistent texture across scans, caused by imperfect scanning (e.g., misalignment between images) or changes in clothing and hairiness.

Independently, noise in lesion detection is unavoidable in practice [12], including both false positives (detections that are not actual lesions) and false negatives (lesions that are undetected). Therefore, matching methods need to not only identify correspondences between of inlier lesions on the source/target (i.e., lesions that are successfully detected in both scans) but also provide the corresponding location on the target/source for lesions that cannot be matched. This additional location information allows physicians to verify whether an unmatchable lesion represents a new growth or is a false positive resulting from noise.

The public dataset introduced by Zhao *et al.* [12] represents a pioneering effort in lesion tracking using 3D meshes. However, the size of the dataset is limited, comprising only 10 subjects. Furthermore, on average only 20 lesion pairs are annotated per subject. The sparse annotation of skin lesions makes the lesion-tracking evaluation far from representative of real-world scenarios.

We propose a framework to match lesions in the context of total body photography using 3D textured meshes while providing locations for unmatchable lesions. To achieve this, we compute accurate correspondence maps relying on the signal represented by the TBP images, in addition to the geometry of the meshes themselves. Using the geometry of the meshes, we first compute coarse correspondence maps taking the source and target meshes to a template mesh. Then, using the lesion/texture signals, we solve for a flow field on the template mesh which is used to refine the correspondence maps. Finally, lesion assignment is performed using the refined correspondence maps. The code will be made publicly available upon acceptance.

We also extend the annotations on the 3DBodyTex dataset [15] to a 25K lesion pairs dataset for skin lesion tracking. To the best of our knowledge, we are the first to release a dataset for skin lesion tracking dataset at this scale. The dataset is available at `https://github.com/weilunhuang-jhu/LesionTrackingDatasetTBP3D`. The dataset includes subjects scanned within a single session under different poses. While the extended dataset does not represent true temporal progression for longitudinal tracking as the ultimate clinical goal, the non-isometric deformation induced by a change in body pose is an issue shared with a change in body shape over time. Non-isometric deformation and sensitivity to texture inconsistencies are immediate technical bottlenecks for finding lesion correspondence in 3D TBP.

Overall, we make three main contributions:

- We propose a novel framework for lesion tracking that automatically matches inlier lesions while providing locations for unmatchable lesions in the context of TBP using 3D textured mesh. The novelty lies in the synergetic integration of geometry- and signal-based refinement via flow-field tailored for the 3D TBP correspondence task.

- We extend the 3DBodyTex dataset by annotating 25K lesion pairs over 198 subjects for skin lesion tracking in 3D TBP under pose variation.
- We validate that the proposed framework outperforms the state-of-the-art methods in both the matching accuracy and the accuracy of the correspondence maps for pose-induced correspondence in 3D TBP. The framework is also more robust to inconsistency between source and target texture, as well as to non-isometric deformations. It exhibits superior accuracy when errors are present in lesion detection.

2. Related work

2.1. Shape correspondence for humans

Shape correspondence between non-rigid surfaces represented as triangle meshes has been an active research topic in computer vision and computer graphics [16, 17, 18]. The shape correspondence problem for triangle meshes is finding a set of corresponding points between two meshes. For human shapes, priors of the human body are commonly applied, such as near-isometric deformation and local rigidity for limbs [19, 20, 21, 22].

Template-based. A line of research relies on a template model, such as SMPL [23], for establishing correspondences across shapes [24, 25, 26]. These methods usually rely on an initial estimation of the pose (body joint positions and orientation) mapping the template to the input. Groueix *et al.* [25] proposed to deform a template mesh to various body poses and shapes with auto-encoder frameworks. Bhatnagar *et al.* [26] used self-supervised learning to register scans of humans to a common 3D human model.

Canonical embedding. Another common shape correspondence method maps vertices into a pose-invariant feature space, where correspondences between the input and template geometries are more easily established [27, 28, 29, 30]. In this category, several shape descriptors have been proposed, from traditional hand-crafted descriptors [31, 32, 33] to deep-learning-based descriptors

[34, 35, 36, 37]. Furthermore, functional map [27] is commonly used for robust regularization.

However, for matching skin lesions, the correspondence map relying on the geometry is not sufficiently accurate. Therefore, using the coarse correspondence map may fail to pair up lesions, particularly if the subject undergoes non-isometric deformation from scan to scan with numerous lesions in close vicinity. We propose leveraging additional signals on the mesh to refine the correspondence map for more accurate matching.

2.2. Graph matching

Given a set of lesions detected in a mesh, we can construct a graph in which a node corresponds to a single lesion and an edge is a connection between a pair of lesions. The node attribute is the position of a lesion, and the edge attribute is the geodesic distance between a pair of lesions. Then, the problem of matching source and target lesions can be formulated as a (partial) graph-matching problem that maximizes the node-to-node and edge-to-edge affinity of the two graphs. Two-graph matching can be modeled as the quadratic assignment problem (QAP), and is known to be NP-hard.

Traditional approaches [38, 39, 40] aim to match graphs by maximizing quadratic objective functions. While effective for simple cases, these methods often struggle with complex graph structures. Some proposed approaches utilize relaxation strategies in graph matching to mitigate the hard combinatorial problem [41, 42, 43]. More recent methods explore hyper-graph matching represented by a tensor to encode the higher-order information, offering increased expressiveness but at the cost of higher computational complexity [44, 45, 46].

Learning-based methods have been shown to improve matching accuracy [47, 48, 49]. Wang *et al.* [47] developed a QAP network to solve the matching problem as a vertex classification task over the association graph whose nodes represent candidate correspondences between the two graphs and edge weights are induced by the affinity matrix built with the two graphs. Liao *et al.* [49] converted the problem of hypergraph matching into a node classification problem and developed a hypergraph neural network. Despite their promise, these methods often require exhaustively annotated datasets for training and struggle to generalize across different domains or datasets. Meanwhile, to address real-world scenarios involving noisy or incomplete data, some techniques are developed for partial graph matching and soft assignment [50, 51].

Despite the advances in graph matching, the solution itself (unlike a correspondence map) does not provide location information for unmatched lesions, an issues that is critical in clinical settings. Starting from coarse correspondence maps, we propose a framework that performs lesion assignment and maps lesions onto a template mesh that allows clinical verification.

2.3. Skin lesion tracking in total body photography

Several works have been proposed for tackling the skin lesion tracking problem over the full body [52, 53, 54, 55]. Korotkov *et al.* [53] designed a TBP system with 21 high-resolution cameras and a turntable to track lesions. However, their method assumes the patient poses are the same across visits and heavily relies on calibrated camera poses for finding lesion correspondence. The work of Korotkov *et al.* [52] improved the earlier method but does not extend to multiple visits. Strzelecki *et al.* [55] developed a TBP system with one digital camera rotating and moving vertically around the subject. The camera captures 32 images for lesion detection and lesion matching [54] based on feature matching and triangulation. However, their method fails when the skin surface is inclined at an angle deviating significantly from 90° with respect to the camera viewing direction. Overall, these methods are limited to a controlled environment and sensitive to camera perspectives and changes in body poses [14].

Recently, the concept of finding lesion correspondence using a 3D representation of the human body has been explored in [11, 12, 13, 14]. Zhao *et al.* [12], Bogo *et al.* [11], and Ahmedt-Aristizabal *et al.* [13] proposed to use a template mesh and rely on anatomical position defined on the template mesh for lesion matching. However, using the template-based correspondence map is insufficient to accurately match lesions. Furthermore, these methods have limited positional expression for lesions since they "snap" the location of a lesion to the nearest vertex. Huang *et al.* [14] proposed to improve the lesion correspondence localization accuracy using landmark-based correspondences refined by texture information. However, their method requires manual annotation of landmarks, is limited to the resolution of the mesh, and is sensitive to inconsistent texture between scans due to scanning artifacts. In this paper, we represent lesion positions using barycentric coordinates within a triangle, making our approach less sensitive to the resolution of the mesh and allowing us to achieve a higher matching accuracy. Additionally, we utilize lesion signals that are agnostic to the textured mesh assuming lesions are provided separately.

Table 1: Summary of notations

$\mathcal{M}_0$	A source mesh in $\mathbb{R}^3$
$\mathcal{M}_1$	A target mesh in $\mathbb{R}^3$
$\mathcal{M}_\mathcal{T}$	A template mesh in $\mathbb{R}^3$
$\mathcal{V}_i$	A set of vertices, $\mathcal{V}_i \subset \mathcal{M}_i$
X_0, X_1	A set of source/target lesions, $X_0 \subset \mathcal{M}_0$, $X_1 \subset \mathcal{M}_1$
ϕ_i^j	A mapping from $\mathcal{M}_i$ to $\mathcal{M}_j$
π	A matching matrix of source and target lesions, $\pi_{ij} = \{0, 1\}$
t^p	The triangle containing p in a mesh
$(t^p, \{\alpha^p, \beta^p, \gamma^p\})$	The barycentric encoding of a point p in a mesh
$\mathcal{F}(t^p)$	The list of vertices of triangle t^p
Φ_i^j	A vertex-to-surface-point mapping, $\Phi_i^j : \mathcal{V}_i \to \mathcal{M}_j$
F_i	A signal on mesh $\mathcal{M}_i$, $F_i : \mathcal{M}_i \to \mathbb{R}$
$\mathcal{I}_0, \mathcal{I}_1$	Source/target texture signals on the template mesh
$\mathcal{L}_0, \mathcal{L}_1$	Source/target lesion signals on the template mesh
$\vec{v}$	A flow field (tangent vector field) on the template mesh
$\exp_p$	The exponential map of a point in a mesh
$D_\mathcal{T}$	The geodesic distance function on the template mesh
$E(\vec{v})$	The energy function of the flow field $\vec{v}$
$E_{X_0,X_1}(\pi)$	The energy function of the matching matrix π

3. Methods

3.1. Method Overview

To track skin lesions in 3D total body photography, we propose using 3D textured meshes for source and target scans. Given the detected lesions in the source/target meshes, we start by establishing coarse correspondence maps from the source/target to a template mesh. We use these maps to define source/target signals over the template domain and then construct a flow field aligning the mapped signals. The coarse correspondence maps for source/target are then refined by advecting forward/backward along the flow field. Finally, lesion assignment is performed using the refined correspondence maps to match lesions and identify unmatchable lesions. An overview of the proposed framework is shown as a block diagram in Fig.1. A summary

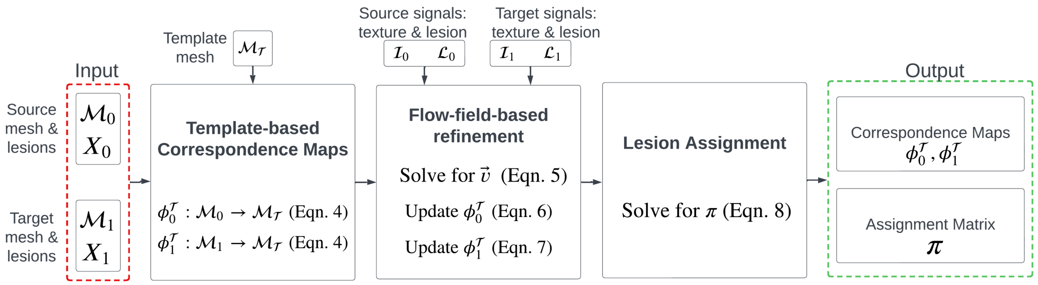


Figure 1: Block diagram of the proposed framework to track skin lesions in 3D total body photography.

of the mathematical notation used in this paper is provided in Table 1.

3.2. Problem formulation

Given a template mesh $\mathcal{M}_\mathcal{T}$, source and target meshes $\mathcal{M}_0$, $\mathcal{M}_1$, and two sets of detected lesions $X_0 \subset \mathcal{M}_0$, $X_1 \subset \mathcal{M}_1$, we would like to find correspondence maps $\phi_0^\mathcal{T} : \mathcal{M}_0 \to \mathcal{M}_\mathcal{T}$ and $\phi_1^\mathcal{T} : \mathcal{M}_1 \to \mathcal{M}_\mathcal{T}$, and a matching matrix $\pi = \{0,1\}^{(|X_0|+1)\times(|X_1|+1)}$ minimizing an energy consisting of two terms:

$$E_{X_0,X_1}(\phi_0^\mathcal{T}, \phi_1^\mathcal{T}, \pi) = \sum_{X_0,X_1} E_{\text{DistanceProximity}}(\phi_0^\mathcal{T}, \phi_1^\mathcal{T}, \pi) + E_{\text{Stochasticity}}(\pi) \ . \quad (1)$$

That is, we want a pair of corresponding source and target lesions to be close to each other while encouraging the correspondence matrix to be doubly stochastic. By adding a dummy lesion to each of the lesion sets in π, we allow the matching function to account for unmatchable lesions. Specifically, assuming a lesion in the source scan can be matched to at most one lesion in the target scan and vice versa, we also enforce:

$$\sum_{j=0}^{|X_1|} \pi_{i,j} = 1, \quad \forall i = 1, \dots, |X_0| \quad (2)$$

$$\sum_{i=0}^{|X_0|} \pi_{i,j} = 1, \quad \forall j = 1, \dots, |X_1| \quad (3)$$

(where $\pi_{i,0} = 1$ indicates a match between the i-th lesion on the source and the target's dummy lesion). Since the dummy lesions can be matched multiple times, the sums involving the dummy lesions can be greater than 1.

3.3. Coarse correspondence map

3.3.1. Template-based coarse correspondence

We start by constructing a coarse correspondence map between the input source/target and a template mesh. We follow the approach from Marin *et al.* [56] to acquire a deformed template mesh registered to the source/target mesh that allows us to construct the correspondence map. Given an input mesh, they propose a localized neural fields network in which a neural field is dedicated to a local region of body shape to predict the vertex displacement of the template mesh (SMPL [23] model). The parameters of the neural field are then refined using Iterative Closest Point [57] through backpropagation. Then, the updated neural field is utilized to register the SMPL model to the input, followed by a refinement that optimizes Chamfer distance. We denote the method SMPL-NICP. Let $\mathcal{M}_{\mathcal{T}}$ be the template mesh, for an input mesh $\mathcal{M}_i,\ i = \{0, 1\}$, the output from SMPL-NICP is a deformed template mesh (i.e., with the same topology as the original template) whose geometry is registered to that of $\mathcal{M}_i$.

We define a correspondence map $\phi_i^{\mathcal{T}} : \mathcal{M}_i \to \mathcal{M}_{\mathcal{T}}$, by first deforming the template mesh to $\mathcal{M}_i$ and then finding, for every point $p \in \mathcal{M}_i$, the nearest surface point on the deformed template. Similarly, we construct a correspondence map $\phi_{\mathcal{T}}^i : \mathcal{M}_{\mathcal{T}} \to \mathcal{M}_i$ by finding the closest surface point on the input mesh $\mathcal{M}_i$ for each point on the deformed template mesh.

We note that $\phi_i^{\mathcal{T}}$ and $\phi_{\mathcal{T}}^i$ are not inverses of each other since two different points on the source/target can have the same closest point on the deformed template. Fig. 2 illustrates the template mesh in (a), the source and the target meshes in (b), the correspondence maps from the source/target to the template in (c), and the source and the target lesions mapped to the template mesh in (d).

3.3.2. Surface point correspondence map

We allow lesions to be located anywhere on the surface on the mesh (i.e., not restricted to vertex positions, a limitation in previous work [12, 13, 14]). To this end, we use barycentric coordinates to encode a point on the mesh: $p \in \mathcal{M} \leftrightarrow (t^p, \{\alpha^p, \beta^p, \gamma^p\})$ where t^p indexes the triangle containing p and $\{\alpha^p, \beta^p, \gamma^p\}$ are the barycentric coordinates of p inside the triangle ($0 \leq \alpha^p, \beta^p, \gamma^p \leq 1$ and $\alpha^p + \beta^p + \gamma^p = 1$).

Using this encoding, we represent mesh correspondences as vertex-to-surface-point maps, taking the vertices on one mesh to points on the second mesh: $\Phi_i^j : \mathcal{V}_i \to \mathcal{M}_j$ is represented by a $\mathbb{R}^{|\mathcal{V}_i| \times 4}$ matrix. The l^{th} row in Φ_i^j maps the l^{th} vertex $v_i^l \in \mathcal{V}_i$ to a point in $\mathcal{M}_j$ in the barycentric encoding.

Given a vertex-to-surface-point correspondence map $\Phi_i^j : \mathcal{V}_i \to \mathcal{M}_j$, we use the barycentric encoding to extend it to a surface-point-to-surface-point correspondence map $\phi_i^j : \mathcal{M}_i \to \mathcal{M}_j$. Concretely, to find the correspondence of a surface point $p \in \mathcal{M}_i$ to the mesh $\mathcal{M}_j$, we map the three vertices of the triangle containing the point p onto the mesh $\mathcal{M}_j$, interpolate the positions of the imaged vertices using the barycentric coordinate of p, and then find the point on $\mathcal{M}_j$ closest to the interpolant. Formally, for a point $p \leftrightarrow (t^p, \{\alpha^p, \beta^p, \gamma^p\}) \in \mathcal{M}_i$ with $\mathcal{F}_i(t^p) = (v_0^p, v_1^p, v_2^p)$ representing the triangle containing p, we have:

$$\phi_i^j(p) = \underset{q \in \mathcal{M}_j}{\arg\min} \, ||\alpha^p \cdot \Phi_i^j(v_0^p) + \beta^p \cdot \Phi_i^j(v_1^p) + \gamma^p \cdot \Phi_i^j(v_2^p) - q|| \, . \tag{4}$$

With Eqn. 4, an arbitrary surface point in the source/target mesh can be mapped to a surface point on the template mesh. The representation of a query point and its mapped point is not restricted to vertex positions. In practice, this operation is accurate if the interpolant is close to the mapped surface point.

3.4. Flow-field-based refinement

The template-based correspondence maps are coarse for two reasons. First, when fitting a template mesh to source/target scan, non-isometric deformation is present at locations near body joints and locations of soft tissues. Second, misalignment between the deformed template mesh and the input mesh occurs if the body pose of the input mesh is far from the canonical “T” pose. Since the coarse correspondence map relies on the nearest point on the registered template mesh to the query point, such a misalignment degrades the accuracy of the mapping. Consequently, a pair of corresponding points in the source and the target will not map to the same position on the template mesh. To refine the correspondence map, we transfer the texture and lesion signals of the source/target to the template mesh using the source/target-to-template correspondences. We then construct a vector field on the template mesh that aligns the transferred signals.

3.4.1. Signal construction on template mesh

Let $F_i : \mathcal{M}_i \to \mathbb{R}$ be a signal on mesh $\mathcal{M}_i$, we transfer the signal to the template mesh using the correspondence map, to define a signal on the template $F_i^{\mathcal{T}} \equiv F_i \circ \phi_{\mathcal{T}}^i : \mathcal{M}_{\mathcal{T}} \to \mathbb{R}$. We consider two types of input signals:

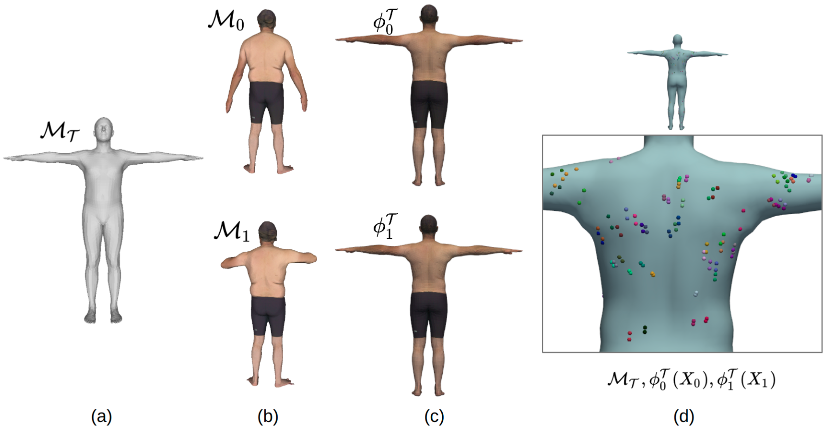


Figure 2: Visualization of the source and target lesions mapped to a template mesh by registering the template mesh to the source and target meshes. (a) shows the template mesh $\mathcal{M}_\mathcal{T}$. (b) shows the source mesh ($\mathcal{M}_0$) and the target mesh ($\mathcal{M}_1$). (c) shows the correspondence maps from the source/target to the template. (d) shows the source and the target lesions mapped to the template mesh. Lesions in correspondences are visualized in the same color.

Texture signal. We construct a triplet of color signals on the template mesh, using the colors in the texture map acquired by the TBP, $\mathcal{I}_0^c, \mathcal{I}_1^c : \mathcal{M}_\mathcal{T} \to \mathbb{R}$, with $c \in \{R, G, B\}$.

Lesion signal. We construct lesion signals on the template mesh using lesion signals defined on the source/target meshes, $\mathcal{L}_0, \mathcal{L}_1 : \mathcal{M}_\mathcal{T} \to \mathbb{R}$. The source/target lesion signals represent the likelihood of a surface point being a lesion. To create the lesion signal, we diffuse a sum of delta functions centered at the lesion positions and normalize the signal across the surface with the maximum signal value to create a scalar-per-vertex signal. In practice, this is done by transferring the lesion positions onto the template mesh and then diffusing over the template mesh.

3.4.2. Surface optical flow

We are given a template mesh $\mathcal{M}_\mathcal{T}$, source/target texture signals $\mathcal{I}_0^c, \mathcal{I}_1^c$, and source/target lesion signals $\mathcal{L}_0, \mathcal{L}_1$. Our goal is to define a tangent vector field $\vec{v}$ on the template mesh such that advection along the field best aligns the source and target signals. To this end we leverage the approach of Prada *et al.* [58], defining the flow field $\vec{v}$ as the minimizer of the energy:

$$\begin{aligned} E(\vec{v}) = &\underbrace{w_{\mathcal{I}} \cdot \sum_{\substack{i\in\{0,1\}\\ c\in\{R,G,B\}}} \int_{\mathcal{M}_{\mathcal{T}}} \Big(\langle \nabla \mathcal{I}_i^c, \vec{v} \rangle - (\mathcal{I}_0^c - \mathcal{I}_1) \Big)^2 \, dp}_{\text{texture fitting}} \\ &+ \underbrace{w_{\mathcal{L}} \cdot \sum_{i=0}^{1} \int_{\mathcal{M}_{\mathcal{T}}} \Big(\langle \nabla \mathcal{L}_i, \vec{v} \rangle - (\mathcal{L}_0 - \mathcal{L}_1) \Big)^2 \, dp}_{\text{lesion fitting}} \\ &+ \underbrace{\epsilon \cdot \int_{\mathcal{M}_{\mathcal{T}}} \| \nabla \vec{v}(p) \|^2 \, dp}_{\text{smoothness}} + \underbrace{\varepsilon \cdot \int_{\mathcal{M}_{\mathcal{T}}} \| \vec{v}(p) \|^2 \, dp}_{\text{size}} \end{aligned} \tag{5}$$

with the first and the second terms penalizing the failure of the vector field to explain the difference in the texture signal and the lesion signal respectively, the third term encouraging the smoothness of the flow, and the fourth term regularizing the norm of the flow to respect the initial correspondence map. We follow the approach proposed by Prada *et al.*, solving for the flow field $\vec{v}$ hierarchically. Please refer to [58] for more details. Fig. 3 visualizes the source and the target (a) texture and (b) lesion signals transferred to the template mesh. Fig. 3 (c) shows the vector field obtained by minimizing Eqn.5.

3.4.3. Update of correspondence map

With the vector field $\vec{v}$ defined on $\mathcal{M}_{\mathcal{T}}$, we update the correspondence map $\phi_0^{\mathcal{T}}$ and $\phi_1^{\mathcal{T}}$ by advecting the positions of correspondence forward and backward along the vector field halfway, separately. Formally, we have:

$$\phi_0^{\mathcal{T}}(p) \leftarrow \exp_{\phi_0^{\mathcal{T}}(p)} \frac{\vec{v}\big(\phi_0^{\mathcal{T}}(p)\big)}{2}, \quad \forall p \in \mathcal{M}_0 \tag{6}$$

and

$$\phi_1^{\mathcal{T}}(p) \leftarrow \exp_{\phi_1^{\mathcal{T}}(p)} \frac{-\vec{v}\big(\phi_1^{\mathcal{T}}(p)\big)}{2}, \quad \forall p \in \mathcal{M}_1 \tag{7}$$

with $\exp_p : T_p\mathcal{M}_{\mathcal{T}} \to \mathcal{M}_{\mathcal{T}}$ the exponential map taking vectors in the tangent space at $p \in \mathcal{M}_{\mathcal{T}}$ to positions on $\mathcal{M}_{\mathcal{T}}$. Through the update of the correspondence maps, the mapped position of a source lesion is expected to be closer to the mapped position of the corresponding target lesion on the template

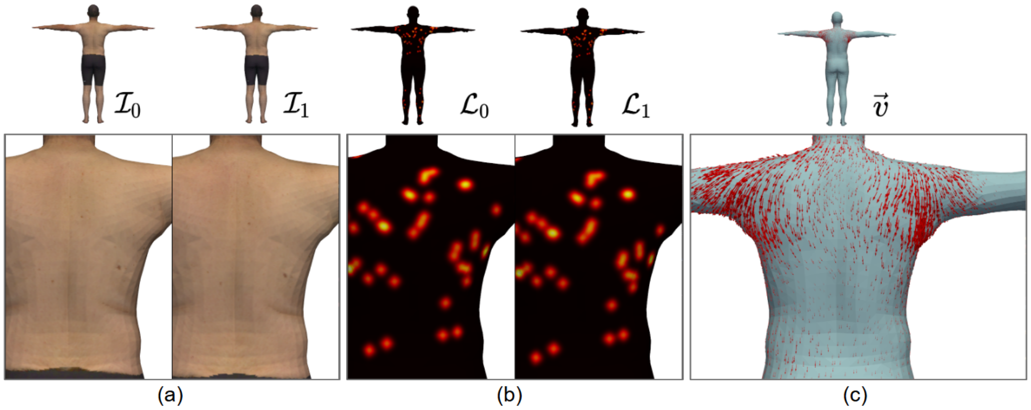


Figure 3: Visualization of the signals on a template mesh and the vector field. (a) shows the source and the target texture signals. (b) shows the source and the target lesion signals. (c) visualizes a solved vector field that explains the difference between the source and the target signals while being smooth and small.

mesh (ideally, at the same position). This is an important step enabling us to match lesions more accurately, particularly when multiple source/target lesions are mapped close to each other by the coarse correspondence maps.

3.5. Lesion assignment

Given source/target correspondence maps $\phi_i^{\mathcal{T}} : \mathcal{M}_i \to \mathcal{M}_{\mathcal{T}}$, we expect lesions, $x_0 \in X_0$ and $x_1 \in X_1$ to be in correspondence if the geodesic distance between $\phi_0^{\mathcal{T}}(x_0)$ and $\phi_1^{\mathcal{T}}(x_1)$ is small. Conversely, we expect $x_0 \in X_0$ (resp. $x_1 \in X_1$) to be unmatched if the geodesic distance from $\phi_0^{\mathcal{T}}(x_0)$ to $\phi_1^{\mathcal{T}}(x_1)$ for all $x_1 \in X_1$ (resp. from $\phi_1^{\mathcal{T}}(x_1)$ to $\phi_0^{\mathcal{T}}(x_0)$ for all $x_0 \in X_0$) is large. We formalize these observations, expressing the assignment matrix $\pi \in \{0,1\}^{(|X_0|+1)\times(|X_1|+1)}$ as the minimizer of the energy:

$$E_{X_0,X_1}(\pi) = \alpha \sum_{x_0 \in X_0, x_1 \in X_1} \pi(x_0,x_1) \cdot D_{\mathcal{T}}(\phi_0^{\mathcal{T}}(x_0), \phi_1^{\mathcal{T}}(x_1)) + \beta\Big[\sum_{x_0 \in X_0} \pi(x_0, x_1^{|X_1|+1}) + \sum_{x_1 \in X_1} \pi(x_0^{|X_0|+1}, x_1)\Big] \quad (8)$$

where $D_{\mathcal{T}} : \mathcal{M}_{\mathcal{T}} \times \mathcal{M}_{\mathcal{T}} \to \mathbb{R}^{\geq 0}$ is the geodesic distance function on $\mathcal{M}_{\mathcal{T}}$. The first term penalizes lesions that are in correspondence but distant from each other, while the second and the third term add a constant cost (β)

to unmatchable lesions in both source and target. The assignment problem can be optimized through the Kuhn–Munkres algorithms [59]. We follow the implementation in Pygmtools [60] to solve for the minimizer of Eqn. 8.

4. Evaluation

4.1. Dataset

4.1.1. Annotation protocol

We extend the 3DBodyTex dataset [15, 61] by annotating lesion correspondence in every pair of meshes. Following the suggestion from a medical expert reported in [62], the lesion identification, while subjective to non-experts annotators (i.e., non-dermatologists), is inclusive of anything that could be potentially considered a skin lesion (e.g., including freckles). The dataset is labeled by four technical annotators (three co-authors, WH, MX, ZL, and one external annotator) using the point list picking function in CloudCompare [63]. A tutorial video was made by WH and shared with all annotators to ensure standardized annotation procedure and results. A standard annotation process for a subject involves loading the two textured meshes in CloudCompare, navigating the 3D view using zoom/pan/rotate operations to select lesions at their centroids, and saving the triangle ID and 3D coordinates for the selected lesions in correspondence. We note that the goal of the annotation is not to identify all the lesions on a subject exhaustively, but rather to ensure correct matches and proper annotated locations for lesions in correspondence. It takes an annotator approximately 15 minutes to label one subject with 100 lesion pairs in two poses (meshes). The annotated lesions in correspondence were manually validated by WH. We excluded subjects when fewer than 10 lesions were found.

The 3DBodyTex.v1 dataset is available for use after a license agreement. We strictly adhered to the dataset's license terms. The study is exempt from full IRB review as it involves secondary analysis of de-identified, public data. We describe our annotation protocol, ensuring transparency regarding who performed the annotations and the quality control measures.

4.1.2. Dataset characteristics

The average number of annotated lesions is 129.6 ($\sigma = 88.4$) across 198 subjects, totaling 25,666 lesions. We define the density of the annotated lesions as the number of neighboring lesions within a geodesic distance of 100 mm. Across all subjects, the average density is 8.5 with $\sigma = 6.1$. In addition, the lesions are distributed with 12,846 lesions on the trunk, 3761 lesions on the upper right limb, 3617 lesions on the upper left limb, 2139 lesions on the lower left limb, 2291 lesions on the lower left limb, and 1012 lesions on the head. Overall, numerous skin lesion pairs are annotated with diversity in body shapes, sizes, poses, and anatomical variations. Therefore, the proposed dataset is suitable for the evaluation of skin lesion tracking that approximates real-world scenarios for TBP. The distribution of the lesion annotations can be found in A.9.

4.1.3. Challenging-pose and numerous-lesion subsets

Since one of our contributions addresses lesion tracking with changes in body pose, we analyze the pose difficulty of the dataset using Procrustes-Aligned Mean Per Joint Position Error (PA-MPJPE). Concretely, the PA-MPJPE is computed relative to a canonical "T" pose for each mesh. Recall from §3.3.1 that an SMPL model is registered to an input mesh. We extract joints from the registered SMPL models for both the input pose and the "T" pose, using shared shape parameters. Then we estimate a rigid transformation to align the two sets of joints. We note that we treat the shape parameters from SMPL as the scale term and do not explicitly estimate it, since the input pose may be far from the "T" pose, making scale estimation difficult. Finally, the meshes are grouped into 35 pose categories provided by 3DBodyTex, and the mean PA-MPJPE is computed for each group. The distribution of PA-MPJPE per pose can be found in A.10.

To demonstrate the performance of our method across different scenarios, we group subjects into subsets for specific purposes. First, we define a "challenging-pose" subset that consists of 26 subjects in which one of the two poses belongs to a group whose mean PA-MPJPE is larger than 200 mm. Unless explicitly stated, this subset is separated from the entire dataset and evaluated individually. Second, we define an "entire" subset with 170 subjects, excluding 26 challenging-poses and an additional 2 cases in which SMPL-NICP failed (visualized in supplement A.11), from the 198 subjects. Furthermore, since lesion tracking is most valuable for patients with numerous and dense lesions, we identify another subset of 32 subjects as a "numerous-lesions" set in which subjects are annotated with more than 200 lesions within the "entire" subset.

4.2. Evaluation of correspondence map

To evaluate the quality of the established correspondence maps, we measure the geodesic distance between a pair of source and target lesions mapped on the template mesh. The evaluation is done for the "entire" subset. We report 1) the average geodesic distance across all the annotated lesion pairs (D_{LP}) and 2) the subject-wise geodesic distance (D_{SW}) as the average geodesic distance of the annotated lesion pairs for the individual subject, and then aggregated across all the subjects. To interpret the geodesic distance between a pair of source and target lesions in a clinical application, a pair of lesions is successfully mapped if the distance between them is less than a threshold criterion. Using the threshold criterion of 10 mm (as in [14]), we measure the success rate for each subject as the percentage of the correctly mapped source and target skin lesions over the total number of annotated skin lesion pairs. We report the subject-wise success rate computed on a pair of meshes (for one subject) and averaged across paired meshes.

We compare our method to two baseline methods that rely solely on geometry to provide correspondence maps, SMPL-NICP [56] and DiffusionNet [35]. We note that DiffusionNet [35] is used as a feature extractor to transform each vertex into a higher-order embedding. Therefore, shape correspondence from DiffusionNet [35] belongs to the category of canonical embedding in shape correspondence (described in §2.1), relying on functional maps to compute correspondence between shapes assuming descriptor preservation over the underlying meshes. Table 2 shows the comparison of the established correspondence maps. For both D_{LP} and D_{SW}, the correspondence maps from our method are substantially more accurate compared to the baseline methods. As a result, the proposed method achieves a success rate of 90.1% ($\sigma = 4.6\%$) at the 10-mm criterion, significantly higher than baseline methods. We note that the reported success rate from [14] is only 57% ($\sigma = 14\%$), computed over 10 subjects and totaling around 200 lesions within our dataset.

Fig. 4 (a) shows the distribution of the geodesic distance between all the lesion pairs. Fig. 4 (b) shows the distribution of the subject-wise geodesic distance between lesion pairs. We observe that the proposed method effectively aligns lesion pairs closer and reduces the long-tail distribution as well. To further investigate the improvement of flow field refinement, we compare correspondence maps established by the proposed method and SMPL-NICP for an individual subject. Fig. 5 (a) shows the distribution of the geodesic distances for all the lesion pairs on the subject. In Fig. 5 (b), we observe that the flow field refinement is also effective in aligning lesion pairs that are

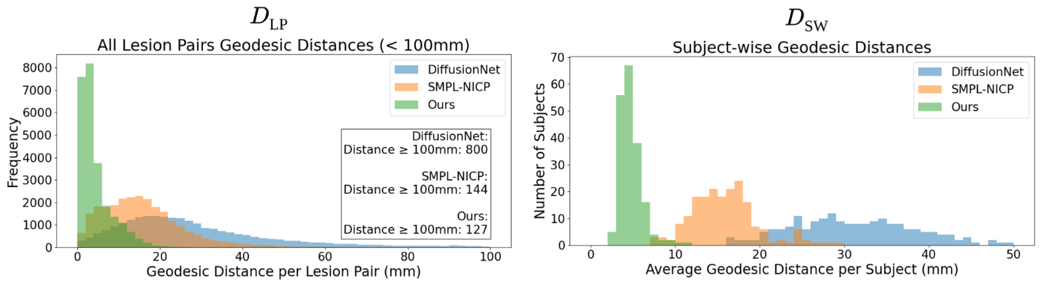


Figure 4: (a) The distribution of all lesion pair geodesic distances. (b) The distribution of the subject-wise geodesic distance between lesion pairs.

originally far from each other (e.g., with a geodesic distance of more than 30 mm). Fig. 5 (c) visualizes the source and the target mesh of the subject, demonstrating challenges of non-isometric deformation due to soft tissue and differing poses. Fig. 5 (d) visualizes the texture signals transferred to the template mesh. In Fig. 5 (e), we show the source and the target lesions mapped onto the template mesh with (ours) and without (SMPL-NICP) flow field refinement. We observe that after the refinement source and target lesions are mapped more closely together, facilitating the task of lesion assignment.

Table 2: Comparison of the accuracy of the established correspondence maps. D_{LP} denotes the geodesic distance across all lesion pairs. D_{SW} represents the geodesic distance between lesion pairs across all the subjects. The standard deviation is shown in brackets.

	DiffusionNet [35]	SMPL-NICP [56]	Ours
D_{LP} (mm)	31.4 (27.4)	16.0 (12.3)	4.9 (8.7)
D_{SW} (mm)	31.6 (8.5)	15.7 (3.9)	4.8 (1.4)
Success rate (%)	12.4 (7.1)	31.7 (14.2)	90.1 (4.6)

4.3. Evaluation of lesion assignment

We compare the proposed framework to existing lesion matching methods from Zhao *et al.* [12] and Ahmedt-Aristizabal *et al.* [13]. We note that the two methods essentially rely on the anatomical position of the source lesions and the target lesions that are mapped in the template mesh while differing in two ways. First, Ahmedt-Aristizabal *et al.* [13] use linear assignment, whereas Zhao *et al.* [12] resort to quadratic assignment using the preservation of geodesic distance between lesion pairs. Second, Ahmedt-Aristizabal *et*

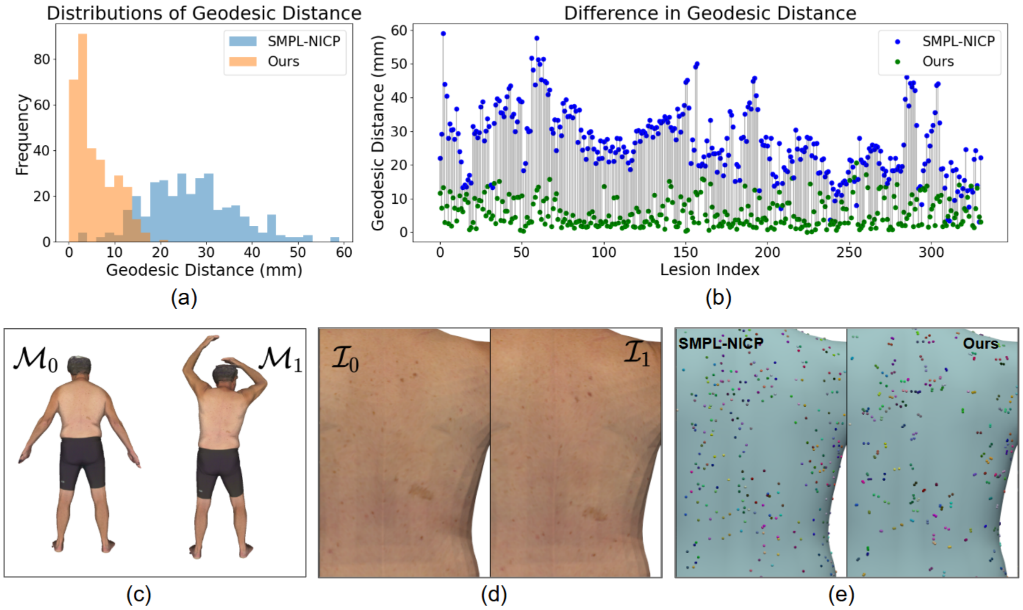


Figure 5: Comparison of correspondence maps for an individual subject (subject-162) between SMPL-NICP and the proposed method. (a) shows the distribution of the geodesic distance for all the lesion pairs on the subject. (b) shows the difference in geodesic distance of each lesion pair between SMPL-NICP and the proposed method. (c) visualizes the source and the target textured meshes. (d) visualizes the source and the target texture signals brought onto the template mesh. (e) visualizes the source and the target lesions mapped to the template mesh using SMPL-NICP and the proposed method. Lesions in correspondences are in the same color.

al. [13] selects LoopReg [26] for template registration, while Zhao *et al.* [12] selects 3D-CODED [25]. To eliminate the effect coming from different registering methods, we re-implement their methods using SMPL-NICP [56] for template registration, similar to what is used in our framework.

For each subject, we calculate the matching accuracy as the number of correctly matched lesions over the number of annotated lesion pairs. Table 3 shows the comparison of the matching accuracy evaluated with different subsets of the dataset. The matching accuracy of the proposed method is highest for all the subsets. In particular, when only evaluating subjects within the "numerous-lesions" subset, we observe that the improvement from our method is more pronounced. In this subset, our method achieves a 98.1% ($\sigma = 1.5\%$) matching accuracy over 32 subjects totaling 8838 lesion pairs, as compared to 91.5% ($\sigma = 17.5\%$) and 95.7% ($\sigma = 5.9\%$) for the methods of Zhao *et al.* [12] and Ahmedt-Aristizabal *et al.* [13], respectively. Therefore, the proposed method effectively pairs up skin lesions for subjects with numerous lesions, the most important benefit of using TBP for skin cancer. The matching accuracy for individual subjects within the "numerous-lesions" subset is shown in A.12. We remark that the reported matching accuracy from Zhao *et al.* [12] is 83% ($\sigma = 38\%$) conducted on 10 subjects and totaling around 200 lesion pairs. We also observe that the proposed method is more robust to differences in topology between source and target meshes. A subject with changes in topology between the two scans and the results can be found in A.13. For the "challenging-pose" subset, our method outperforms the other two baseline approaches from Zhao *et al.* [12] and Ahmedt-Aristizabal *et al.* [13] as well. However, our matching accuracy drops to 95.9% ($\sigma = 9.0\%$) for this subset. We note that the template-based correspondence map could fail when the poses are challenging, either unseen in the training dataset (in SMPL-NICP [56]) or far from the template "T" pose.

4.4. Impact of pose difficulty

To further investigate the impact of pose difficulty on performance, we plot the PA-MPJPE against the matching accuracy. In the 3DBodyTex dataset, subjects were scanned once in either the "U" pose or the "A" pose. These two standard poses are relatively easy to register among the 35 pose categories, especially the "U" pose found to have the lowest PA-MPJPE in A.10). Since subjects were scanned in two poses, we use the other pose (not the "U" pose nor the "A" pose) of each subject for the pose complexity. We perform stratified aggregation on pose complexity. Subjects are first clustered by pose category to derive a per-pose mean PA-MPJPE. These poses are subsequently discretized into quantiles (bins) based on the per-pose mean PA-MPJPE, allowing us to evaluate matching performance as a function of pose difficulty. All 198 subjects (396 scans) are evaluated here.

Fig. 6 (a) shows the average matching accuracy over the mean PA-MPJPE for each bin. We observe a decay in matching accuracy when the mean PA-MPJPE is larger than 180 mm for both the proposed method and the method from Ahmedt-Aristizabal *et al.* [13]. We also plot the lesion pair geodesic distance (D_{LP}) across the template mesh aggregated over all subjects in Fig. 6 (b). We observe that the distances (errors) tend to be larger around regions that are prone to non-isometric deformation and change in mesh connectivity, such as armpits, inner elbow, and belly.

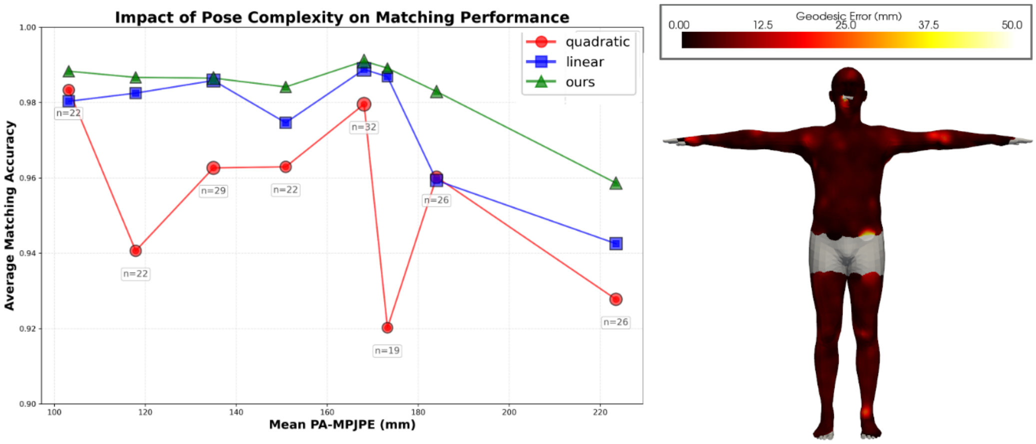
Table 3: Comparison of the matching accuracy (%) across all the subjects within different subsets of the dataset. The standard deviation is shown in brackets. Recall that the "entire" subset has 170 subjects, excluding 26 subjects in the "challenging-pose" subset. There are 32 subjects in the "numerous-lesions" subset.

	Zhao *et al.* [12]	Ahmedt-Aristizabal *et al.* [13]	Ours
entire	96.2 (11.5)	98.3 (3.4)	98.8 (1.7)
numerous-lesions	91.5 (17.5)	95.7 (5.9)	98.1 (1.5)
challenging-pose	92.8 (11.6)	94.3 (10.4)	95.9 (9.0)

4.5. Performance in noisy lesion detection

We evaluate the performance of the proposed framework for noisy lesion detection by independently taking out $p\%$ of lesions in the source and target, and then pairing up lesions. The evaluation is performed on the "numerous-lesions" subset. We compare the proposed framework to the approaches from Zhao *et al.* [12] and Ahmedt-Aristizabal *et al.* [13].

We compute precision, recall, and F1 scores between the source and target lesions for each subject and then report the average and standard deviation across subjects. The F1 score is the harmonic mean of precision and recall. Given predicted matrix π_{pred} containing k_{pred} matches, and ground truth matrix π_{gt} containing k_{gt} matches, denote $\odot$ as element-wise product, we



(a) (b)

Figure 6: Analysis of pose's impact on performance. (a) shows the average matching accuracy over the mean PA-MPJPE. "linear" and "quadratic" represent the methods of Ahmedt-Aristizabal *et al.* [13] and Zhao *et al.* [12], respectively. The number of subjects (n) for each pose complexity is indicated. (b) shows the error distribution (lesion pair geodesic distance) across the template mesh aggregated over all subjects. The gray mask indicates region without any annotated lesion pairs.

have:

$$\text{precision} = \sum(\pi_{pred} \odot \pi_{gt})/k_{pred}, \tag{9}$$

$$\text{recall} = \sum(\pi_{pred} \odot \pi_{gt})/k_{gt}, \tag{10}$$

$$\text{F1} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}. \tag{11}$$

Fig. 7 compares the precision, recall, and F1 scores for the three methods under different noise levels. Compared to the baseline methods, the proposed method consistently maintains the highest F1 score across noise levels smaller than 25%. However, the relative advantage over the linear baseline narrows as noise increases. This is consistent with our ablation study (§4.6.2) showing reliance on the lesion signal; as that signal degrades, the benefit of our signal-driven refinement naturally diminishes. We note that the reported recall for lesion detection by Zhao *et al.* [12] ranges from 78% to 96%, validating that the noise range in our experiment is practical. Furthermore, the quadratic assignment using the preservation of geodesic distance between lesion pairs is known to be sensitive to noise, which is corroborated in our findings.

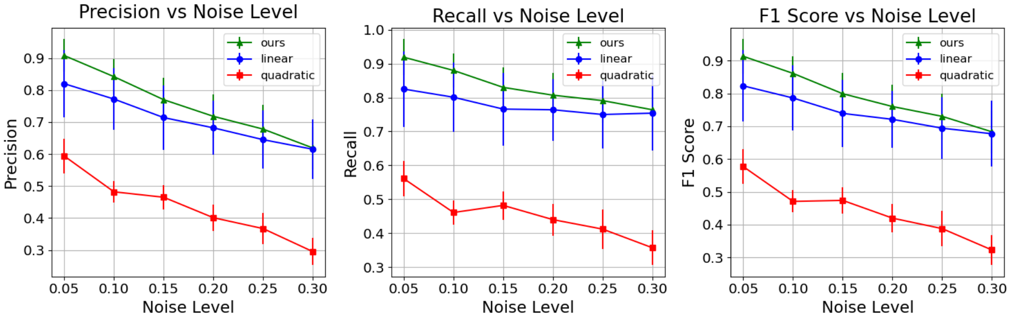


Figure 7: Comparison of the precision, recall, and F1 scores under different noise levels. "linear" and "quadratic" represent the methods of Ahmedt-Aristizabal *et al.* [13] and Zhao *et al.* [12], respectively. The noise level is the percentage of lesions independently taken out from the source and target lesions, ranging from 5% to 30% every 5%.

4.6. Ablation study

4.6.1. Initial correspondence map

Conceptually, the proposed flow field refinement can be applied to any shape correspondence method that maps the source and target mesh to a template mesh. However, the level of improvement might be different, depending on the consistency of the source and target signals constructed on the template mesh. Therefore, we compare how the flow field refinement works on different initial correspondence maps. We evaluate the average geodesic distance across all the annotated lesion pairs (D_{LP}), using DiffusionNet [35] and SMPL-NICP [56] for initial correspondence maps. The D_{LP} for DiffusionNet and SMPL-NICP are 31.4 mm ($\sigma = 27.4$) and 16.0 mm ($\sigma = 12.3$) initially, and 27.4 mm ($\sigma = 18.0$) and 4.9 mm ($\sigma = 8.7$) after refinement, respectively. We show that the proposed flow field refinement effectively improves the correspondence map from both methods. Remarkably, the improvement for a more precise initial correspondence map is more signnificant – an improvement of 11.1 mm for SMPL-NICP as compared to an improvement of 4.0 mm for DiffusionNet. Since SMPL-NICP gives more consistent source and target signals to be aligned, in line with the assumptions of brightness constancy and little motion [64, 65], it benefits more from flow field refinement.

4.6.2. Flow field refinement

Usage of signals. To investigate the effectiveness of different signals to refine the correspondence maps, we compare D_{LP} by using only the texture signal, the lesion signal, and a combination of the two. On the "numerous-lesion" subset, we observe that using the texture signal gives 13.6 mm ($\sigma = 12.0$), while there is no significant difference between using only the lesion signal compared to using a combination of the two, with both giving 5.0 mm ($\sigma = 7.2$). Overall, the signals used in our method refine the correspondence maps while being robust to inconsistent source and target texture coming from scanning artifacts.

Usage of small magnitude regularization. To respect the initial correspondence map and be robust to inconsistency in signals due to noise (e.g., undetected lesions or texture that does not agree in the source and target), we add a magnitude regularization to enforce the vector field to be small. On the "numerous-lesion" subset, we observe that the D_{LP} is 15.9 mm ($\sigma = 11.8$) without the regularization and 5.0 mm ($\sigma = 7.2$) with the regularization.

5. Discussion

5.1. Choice of barycentric coordinate representation

The choice of the barycentric coordinate representation for lesions (and correspondence maps) tackles the limitation of inaccurate and resolution-dependent matching present in previous works [12, 13, 14]. Since the barycentric coordinate representation impacts both the accuracy of lesion location and the initial correspondence map, it significantly affects the quality of signal construction on the template mesh, and consequently downstream evaluation tasks. As such, the barycentric coordinate representation is crucial for our accurate correspondence maps and high matching performance.

As a comparison, on the entire dataset, the average matching accuracy across all the subjects using the vertex representation is 83.5% ($\sigma = 11.1$), as compared to 98.3% ($\sigma = 3.4$) evaluated with our re-implementation of the approach from Ahmedt-Aristizabal *et al.* [13]. As expected, we observe a consistent decrease in matching accuracy as the lesion count increases due to inaccurate representation. A detailed comparison of the average matching accuracy for different numbers of lesions per subject can be found in A.14.

In addition, the selected representation is robust to the input mesh resolution. Our evaluation is conducted on the low-resolution mesh from the

3DBodyTex dataset with 10K vertices on average, whereas both Zhao *et al.* [12] and Huang et al. [14] use high-resolution mesh with 300K vertices from the same dataset. To investigate the effect of input mesh resolution, we perform the flow field refinement using the texture signal on the high-resolution mesh and compare to the results in §4.6.2. On the "numerous-lesion" subset, the improvement of D_{LP} from using high-resolution texture (as compared to low-resolution texture) is only 0.2 mm. The improvement becomes even smaller (0.1 mm) after introducing the lesion signal.

However, the performance of our method depends on the template mesh resolution at run-time. Our implementation uses nearest-sampling to construct the texture signal and diffuses a sum of delta functions centered at mapped lesion positions on the template mesh for lesion signals. Concretely, the template mesh (from SMPL [23]) has 6890 vertices. To ensure sufficient texture/lesion signals on the template mesh, we perform three rounds of one-to-four subdivisions (to 440,834 vertices) on the template mesh when constructing the signals (in §3.4.1) and solving the surface optical flow (in §3.4.2). The number of iterations for one-to-four subdivisions is determined empirically: a lower resolution produces a larger error (in geodesic distance) for lesion pairs, while a higher resolution is slow and does not provide noticeable improvement.

5.2. Computational cost

Overall, our method achieves state-of-the-art matching accuracy without sacrificing efficiency. For each subject, our framework takes around 7 minutes, including the following components: 1) Coarse correspondence map (4 minutes). 2) Signal construction on template mesh (20 seconds). 3) Surface optical flow (150 seconds). 4) Lesion assignment (1 second). In particular, for the "numerous-lesion" subset, our method performs significantly better than the approach from Ahmedt-Aristizabal *et al.* [13] with an acceptable computation overhead of 3 minutes. On the other hand, the quadratic assignment used by Zhao *et al.* [12] is NP-hard.

5.3. Evaluation on subjects with dark skin tones

Our dataset includes a variety of skin tones, notably including 5 subjects with dark skin tones. We observe that the number of skin lesions is relatively small for those subjects, with 26.6 ($\sigma = 7.7$) lesions on average. Our method successfully pairs up all the lesions on the 5 subjects. Example results can be found in A.15.

5.4. Clinical Integration

The proposed method serves as a correspondence machine that could be integrated into an automatic screening system for skin cancer. For example, once a new 3D TBP scan is acquired, existing methods (e.g., [12]) can be applied to detect skin lesions on the scan. Subsequently, our framework establishes lesions in correspondence and detects new/disappearing lesions, creating a visual record of all moles on a person's body. The computed pose difficulty value can suggest to doctors to pay extra attention to the results in difficult pose. Meanwhile, the visual presentation of the registered template mesh and the input mesh allows clinicians to reject failure cases. Finally, downstream tasks can be built on the proposed framework, such as monitoring lesions that have significantly changed. The current computational time, while optimizable, is reasonable to streamline a complete skin examination of the full body.

5.5. Limitations

The proposed framework has several limitations. 1) Our method fails to accurately map the source and the target lesions on the template mesh within the 10-mm criterion if a large non-isometric deformation exists. Fig. 8 (a) visualizes a subject with large non-isometric deformation around the chest and the belly area. Fig. 8 (b) demonstrates the source and the target lesions imaged on the template mesh. Although the proposed flow-field-based refinement brings closer lesions in correspondence, the lesion pairs in the red box are still far from each other, with a geodesic distance of more than 50 mm. However, these lesions are correctly paired up in our lesion assignment step. 2) The proposed framework fails for some challenging poses when the template mesh is incorrectly registered to the input mesh. For example, Fig. 8 (c) shows an example from "challenging-pose" subset where template-based registration fails. The registered template mesh flips the left and the right legs. As a result, the geodesic distance for lesions in correspondence is excessively large on those incorrectly registered body regions. Our method cannot match those lesions successfully. Furthermore, while our pose difficulty metric could be used to warn clinicians for difficult cases, the metric fails to categorize the two failure cases (in A.11) as challenging poses. We believe that it is confusion due to bilateral symmetry that makes template fitting difficult, particularly when limbs cross the body. A pose difficulty metric incorporating symmetry should be investigated in the future. Nonetheless, some challenging poses may not be common in clinical settings, especially

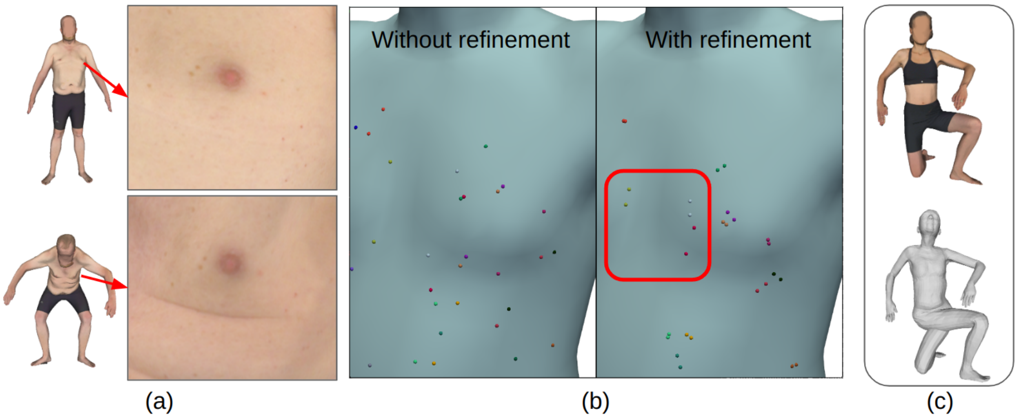


Figure 8: Examples of failure cases. (a) visualizes a subject undergoing large non-isometric deformation. (b) illustrates the source and the target lesions imaged on the template mesh for lesions shown in (a). In the red box, we highlight lesion pairs with a geodesic distance of more than 50 mm. (c) shows a failure in template registration of a pose categorized as "challenging-pose".

considering that patients have to maintain the pose during the scan. 3) The proposed dataset exhibits limited annotations of distal extremities, such as lesions on fingers and toes. In particular, due to complex morphology, these anatomical regions are challenging for digital imaging. Consequently, the proposed framework may demonstrate reduced performance and reliability for lesions in these regions. 4) The 3DBodyTex dataset only contains pose variations within a single session and lacks true longitudinal data (e.g., new/removed lesions, morphological changes). Future work should consider the evaluation of true longitudinal changes.

6. Conclusions

In this paper, we propose a framework to match lesions in the context of full body using 3D textured meshes while providing locations for unmatchable lesions. We propose a skin lesion tracking dataset with 25K lesion pairs over 198 subjects. To our knowledge, a dataset for skin lesion tracking in 3D TBP at this scale is not reported in the literature.. We show that the proposed method effectively refines correspondence maps to align lesion pairs, achieving a success rate of 90.1% at 10-mm criterion. The proposed framework accomplishes state-of-the-art matching accuracy, an accuracy of 98.1%

for 32 subjects with more than 200 lesions on the body. Furthermore, our method is validated to be more robust to inconsistent texture between the source and the target meshes and outperforms state-of-the-art in noisy lesion detection.

In the future, we would like to extend the framework for more than two TBP scans. With more than two TBP scans, some of the false positives and false negatives may potentially be resolved by evaluating the consistency of a lesion's life cycle [66]. Moreover, the method needs to be evaluated on longitudinal data with a longer duration that may include more significant changes in skin conditions.

7. Acknowledgments

The research was in part supported by the Intramural Research Program (IRP) of the NIH/NICHD, Phase I of NSF STTR grant 2127051, Phase II of NSF STTR grant 2335086, and Phase I NIH/NIBIB STTR grant R41EB032304. We thank Ryan Whittaker for the help with annotating skin lesion pairs.

8. Declaration of generative AI and AI-assisted technologies in the writing process

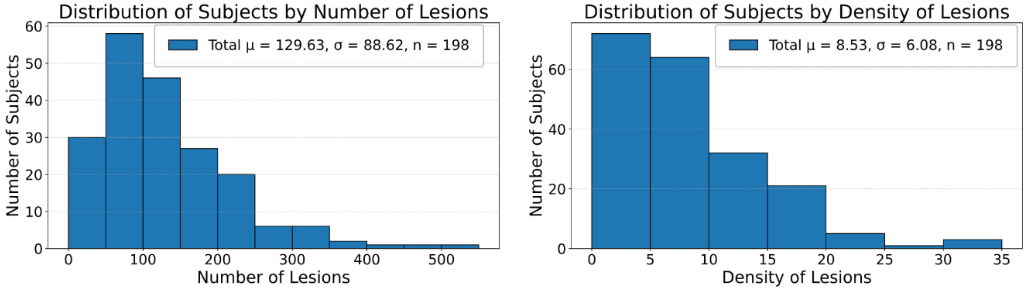
During the preparation of this work, the authors used Claude AI and Google Gemini in order to polish the wording. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Appendix A.

References

[1] N. R. Abbasi, H. M. Shaw, D. S. Rigel, R. J. Friedman, W. H. McCarthy, I. Osman, A. W. Kopf, D. Polsky, Early diagnosis of cutaneous melanoma: revisiting the abcd criteria, Jama 292 (22) (2004) 2771–2776.

[2] A. C. Halpern, Total body skin imaging as an aid to melanoma detection., in: Seminars in cutaneous medicine and surgery, Vol. 22, 2003, pp. 2–8.



(a) (b)

Figure A.9: (a) The distribution of subjects by the number of annotated lesions. (b) The distribution of subjects by the density of annotated lesions. The density of the annotated lesions is defined as the number of neighboring lesions within a geodesic distance of 100 mm.

[3] J. E. Rayner, A. M. Laino, K. L. Nufer, L. Adams, A. P. Raphael, S. W. Menzies, H. P. Soyer, Clinical perspective of 3d total body photography for early detection and screening of melanoma, Frontiers in Medicine 5 (2018) 152.

[4] K. Grochulska, B. Betz-Stablein, C. Rutjes, F. P.-C. Chiu, S. W. Menzies, H. P. Soyer, M. Janda, The additive value of 3d total body imaging for sequential monitoring of skin lesions: a case series, Dermatology 238 (1) (2022) 12–17.

[5] T. Deinlein, C. Michor, R. Hofmann-Wellenhof, K. Schmid-Zalaudek, R. Fink-Puches, The importance of total-body photography and sequential digital dermatoscopy for monitoring patients at increased melanoma risk, JDDG: Journal der Deutschen Dermatologischen Gesellschaft 18 (7) (2020) 692–697.

[6] A. Hornung, T. Steeb, A. Wessely, T. J. Brinker, T. Breakell, M. Erdmann, C. Berking, M. V. Heppt, The value of total body photography for the early detection of melanoma: a systematic review, International Journal of Environmental Research and Public Health 18 (4) (2021) 1726.

[7] A. Ji-Xu, J. Dinnes, R. Matin, Total body photography for the diagnosis of cutaneous melanoma in adults: a systematic review and meta-analysis, British Journal of Dermatology 185 (2) (2021) 302–312.

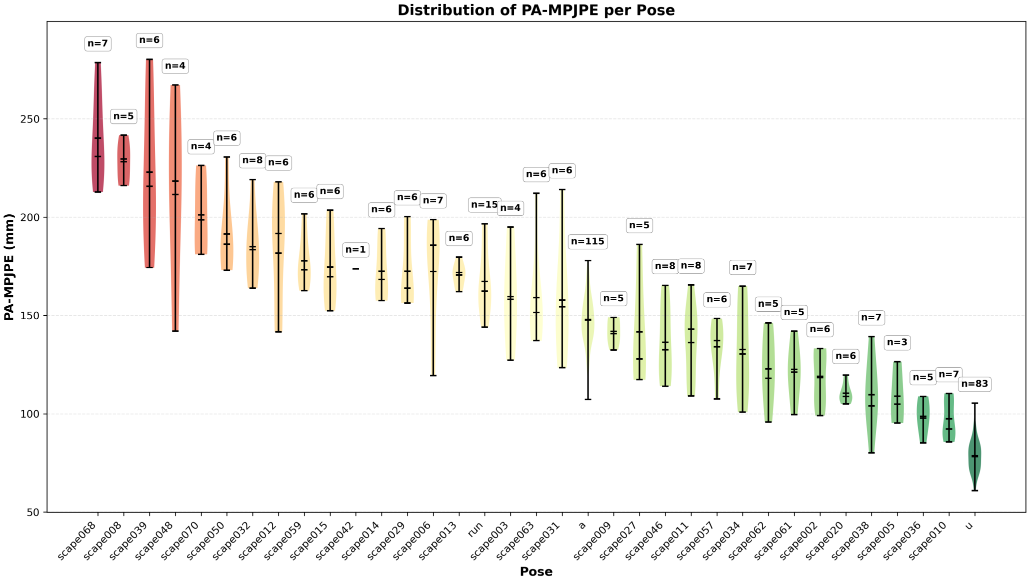


Figure A.10: The distribution of Procrustes-Aligned Mean Per Joint Position Error (PA-MPJPE) for 35 pose categories used in the 3DBodyTex dataset. The number of subjects (n) for each pose category is indicated.

[8] N. Guido, E. L. Hagstrom, E. Ibler, C. Carneiro, K. A. Orrell, R. C. Kelm, A. W. Rademaker, D. P. West, B. Nardone, A novel total body digital photography smartphone application designed to detect and monitor skin lesions: a pilot study, Journal of Surgical Dermatology 6 (2) (2021) 14–18.

[9] C. A. Primiero, A. M. McInerney-Leo, B. Betz-Stablein, D. C. Whiteman, L. Gordon, L. Caffery, J. F. Aitken, E. Eakin, S. Osborne, L. Gray, et al., Evaluation of the efficacy of 3d total-body photography with sequential digital dermoscopy in a high-risk melanoma cohort: protocol for a randomised controlled trial, BMJ open 9 (11) (2019) e032969.

[10] C. A. Primiero, G. G. Rezze, L. J. Caffery, C. Carrera, S. Podlipnik, N. Espinosa, S. Puig, M. Janda, H. P. Soyer, J. Malvehy, A narrative review: opportunities and challenges in artificial intelligence skin image analyses using total body photography, Journal of Investigative Dermatology (2024).

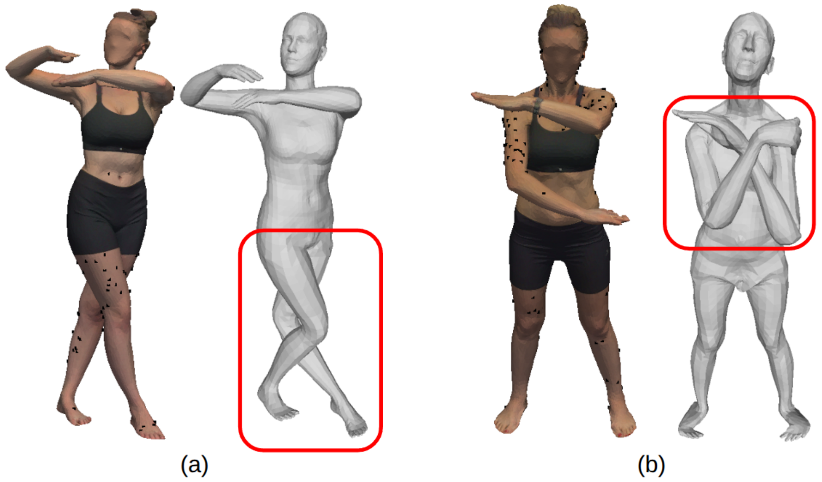


Figure A.11: The two cases in which NICP-SMPL fails to register the template mesh to the input mesh of (a) subject 15 and (b) subject 73. The poorly registered body region is highlighted in the red box. The annotated lesions for the two subjects are visualized as black dots.

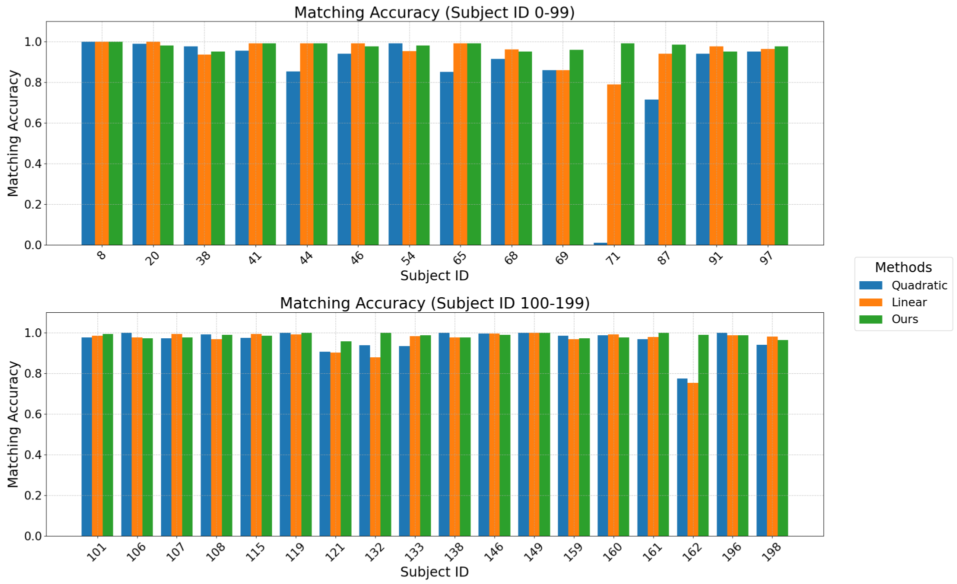


Figure A.12: The matching accuracy for individual subjects within the "numerous-lesions" subset, totaling 32 subjects.

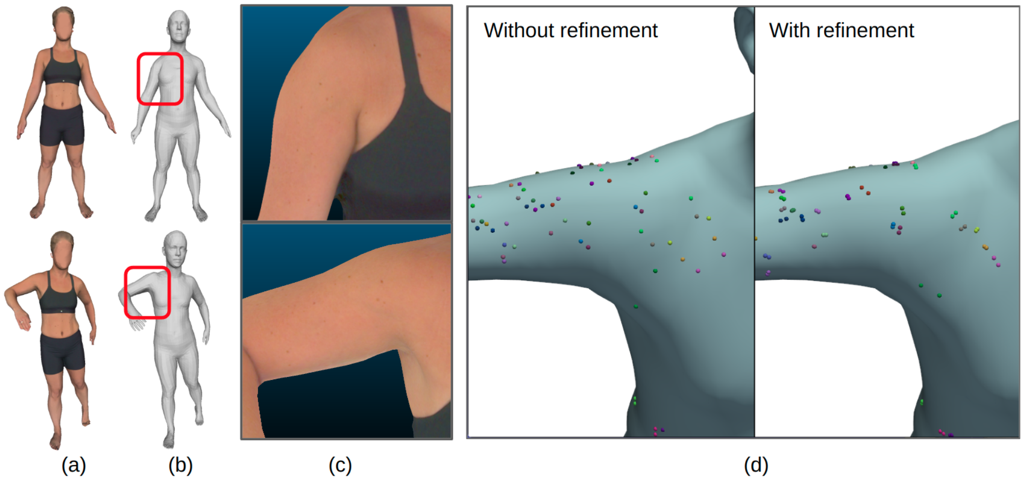


Figure A.13: An example demonstrating that the proposed method is more robust to change in topology. (a) visualizes the source and the target mesh of subject 71. (b) visualized the registered template mesh. Differences in topology between the two registered meshes are highlighted in the red box. (c) shows a zoom-in view of the armpit location. (d) shows the source and the target lesions mapped on the template mesh with and without the vector-field-based refinement. We observe that lesions are aligned closer with the refinement. As a result, lesion pairs are mostly correctly matched (shown in Fig A.12).

[11] F. Bogo, J. Romero, E. Peserico, M. J. Black, Automated detection of new or evolving melanocytic lesions using a 3d body model, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2014, pp. 593–600.

[12] M. Zhao, J. Kawahara, K. Abhishek, S. Shamanian, G. Hamarneh, Skin3d: Detection and longitudinal tracking of pigmented skin lesions in 3d total-body textured meshes, Medical Image Analysis 77 (2022) 102329.

[13] D. Ahmedt-Aristizabal, C. Nguyen, L. Tychsen-Smith, A. Stacey, S. Li, J. Pathikulangara, L. Petersson, D. Wang, Monitoring of pigmented skin lesions using 3d whole body imaging, Computer Methods and Programs in Biomedicine 232 (2023) 107451.

[14] W.-L. Huang, D. Tashayyod, J. Kang, A. Gandjbakhche, M. Kazhdan, M. Armand, Skin lesion correspondence localization in total body pho-

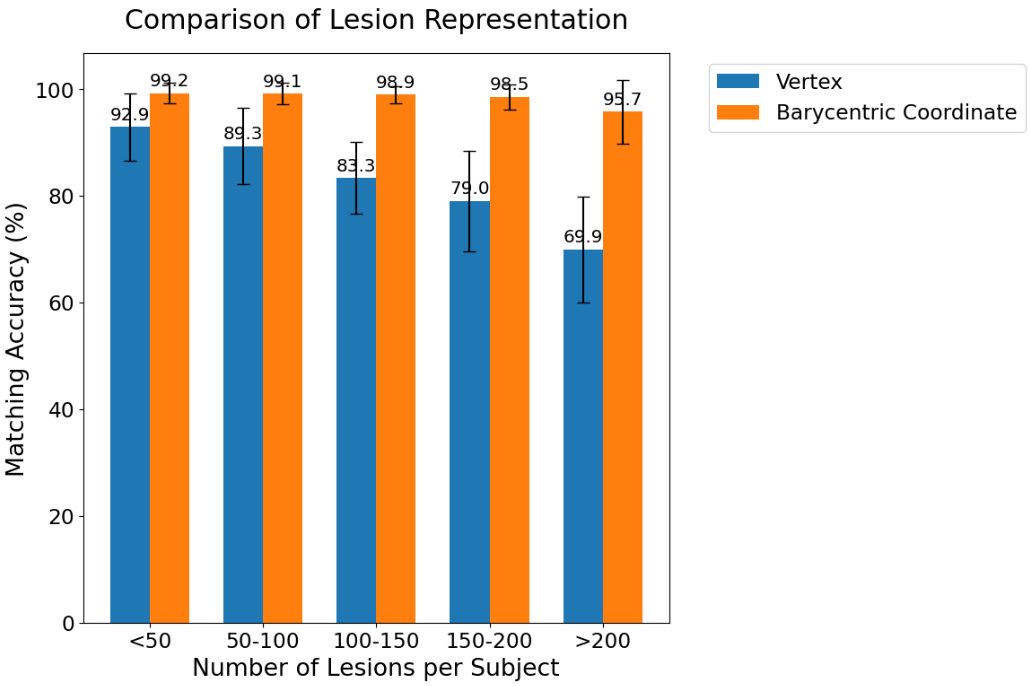


Figure A.14: Comparison of matching accuracy for different lesion representations.

tography, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2023, pp. 260–269.

[15] A. Saint, E. Ahmed, K. Cherenkova, G. Gusev, D. Aouada, B. Ottersten, et al., 3dbodytex: Textured 3d body dataset, in: 2018 International Conference on 3D Vision (3DV), IEEE, 2018, pp. 495–504.

[16] O. Van Kaick, H. Zhang, G. Hamarneh, D. Cohen-Or, A survey on shape correspondence, in: Computer graphics forum, Vol. 30, Wiley Online Library, 2011, pp. 1681–1707.

[17] Y. Sahillioğlu, Recent advances in shape correspondence, The Visual Computer 36 (8) (2020) 1705–1721.

[18] B. Deng, Y. Yao, R. M. Dyke, J. Zhang, A survey of non-rigid 3d registration, in: Computer Graphics Forum, Vol. 41, Wiley Online Library, 2022, pp. 559–589.

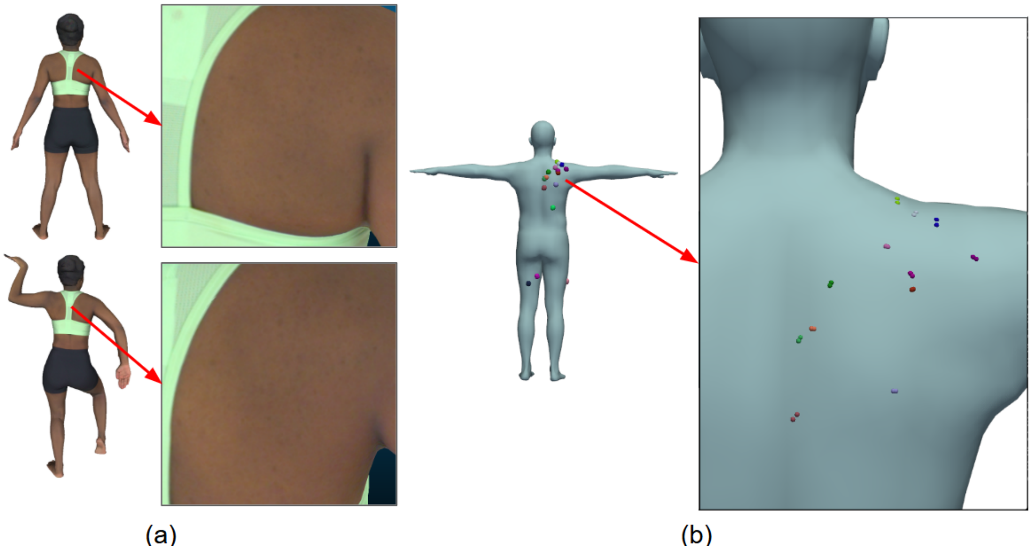


Figure A.15: An example subject with dark skin. (a) visualizes the source and the target mesh for subject 93 with a close-up view of the top-right region of the back with lesions. (b) visualizes the template mesh with the source and target lessons mapped to the template mesh. The subject has 27 annotated lesions, and all of them are successfully matched by the proposed framework.

[19] S. Zuffi, M. J. Black, The stitched puppet: A graphical model of 3d human shape and pose, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3537–3546.

[20] S. Wang, A. Geiger, S. Tang, Locally aware piecewise transformation fields for 3d human mesh registration, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 7639–7648.

[21] B. L. Bhatnagar, C. Sminchisescu, C. Theobalt, G. Pons-Moll, Combining implicit function learning and parametric models for 3d human reconstruction, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16, Springer, 2020, pp. 311–329.

[22] T. Alldieck, H. Xu, C. Sminchisescu, imghum: Implicit generative models of 3d human shape and articulated pose, in: Proceedings of the

IEEE/CVF International Conference on Computer Vision, 2021, pp. 5461–5470.

[23] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, M. J. Black, Smpl: A skinned multi-person linear model, in: Seminal Graphics Papers: Pushing the Boundaries, Volume 2, 2023, pp. 851–866.

[24] X. Huang, H. Yang, E. Vouga, Q. Huang, Dense correspondences between human bodies via learning transformation synchronization on graphs, Advances in Neural Information Processing Systems 33 (2020) 17489–17501.

[25] T. Groueix, M. Fisher, V. G. Kim, B. C. Russell, M. Aubry, 3d-coded: 3d correspondences by deep deformation, in: Proceedings of the european conference on computer vision (ECCV), 2018, pp. 230–246.

[26] B. L. Bhatnagar, C. Sminchisescu, C. Theobalt, G. Pons-Moll, Loopreg: Self-supervised learning of implicit surface correspondences, pose and shape for 3d human mesh registration, Advances in Neural Information Processing Systems 33 (2020) 12909–12922.

[27] M. Ovsjanikov, M. Ben-Chen, J. Solomon, A. Butscher, L. Guibas, Functional maps: a flexible representation of maps between shapes, ACM Transactions on Graphics (ToG) 31 (4) (2012) 1–11.

[28] O. Litany, T. Remez, E. Rodola, A. Bronstein, M. Bronstein, Deep functional maps: Structured prediction for dense shape correspondence, in: Proceedings of the IEEE international conference on computer vision, 2017, pp. 5659–5667.

[29] R. Marin, S. Melzi, E. Rodola, U. Castellani, Farm: Functional automatic registration method for 3d human bodies, in: Computer Graphics Forum, Vol. 39, Wiley Online Library, 2020, pp. 160–173.

[30] N. Donati, A. Sharma, M. Ovsjanikov, Deep geometric functional maps: Robust feature learning for shape correspondence, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 8592–8601.

[31] J. Sun, M. Ovsjanikov, L. Guibas, A concise and provably informative multi-scale signature based on heat diffusion, in: Computer graphics forum, Vol. 28, Wiley Online Library, 2009, pp. 1383–1392.

[32] S. Salti, F. Tombari, L. Di Stefano, Shot: Unique signatures of histograms for surface and texture description, Computer Vision and Image Understanding 125 (2014) 251–264.

[33] T. W. Mitchel, S. Rusinkiewicz, G. S. Chirikjian, M. Kazhdan, Echo: Extended convolution histogram of orientations for local surface description, in: Computer Graphics Forum, Vol. 40, Wiley Online Library, 2021, pp. 180–194.

[34] L. Wei, Q. Huang, D. Ceylan, E. Vouga, H. Li, Dense human body correspondences using convolutional networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 1544–1553.

[35] N. Sharp, S. Attaiki, K. Crane, M. Ovsjanikov, Diffusionnet: Discretization agnostic learning on surfaces, ACM Transactions on Graphics (TOG) 41 (3) (2022) 1–16.

[36] T. W. Mitchel, V. G. Kim, M. Kazhdan, Field convolutions for surface cnns, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 10001–10011.

[37] H. Kim, J. Kim, J. Kam, J. Park, S. Lee, Deep virtual markers for articulated 3d shapes, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 11615–11625.

[38] M. Leordeanu, M. Hebert, R. Sukthankar, An integer projected fixed point method for graph matching and map inference, Advances in neural information processing systems 22 (2009).

[39] O. Enqvist, K. Josephson, F. Kahl, Optimal correspondences from pairwise constraints, in: 2009 IEEE 12th international conference on computer vision, IEEE, 2009, pp. 1295–1302.

[40] M. Cho, J. Lee, K. M. Lee, Reweighted random walks for graph matching, in: Computer Vision–ECCV 2010: 11th European Conference on

Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part V 11, Springer, 2010, pp. 492–505.

[41] M. Leordeanu, M. Hebert, A spectral technique for correspondence problems using pairwise constraints, in: Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, Vol. 2, IEEE, 2005, pp. 1482–1489.

[42] U. Kang, M. Hebert, S. Park, Fast and scalable approximate spectral graph matching for correspondence problems, Information Sciences 220 (2013) 306–318.

[43] C. Schellewald, C. Schnörr, Probabilistic subgraph matching based on convex relaxation, in: International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition, Springer, 2005, pp. 171–186.

[44] O. Duchenne, F. Bach, I.-S. Kweon, J. Ponce, A tensor-based algorithm for high-order graph matching, IEEE transactions on pattern analysis and machine intelligence 33 (12) (2011) 2383–2395.

[45] S. Park, S.-K. Park, M. Hebert, Fast and scalable approximate spectral matching for higher order graph matching, IEEE transactions on pattern analysis and machine intelligence 36 (3) (2013) 479–492.

[46] J. Yan, C. Zhang, H. Zha, W. Liu, X. Yang, S. M. Chu, Discrete hypergraph matching, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1520–1528.

[47] R. Wang, J. Yan, X. Yang, Neural graph matching network: Learning lawler's quadratic assignment problem with extension to hypergraph and multiple-graph matching, IEEE Transactions on Pattern Analysis and Machine Intelligence 44 (9) (2021) 5261–5279.

[48] M. Rolínek, P. Swoboda, D. Zietlow, A. Paulus, V. Musil, G. Martius, Deep graph matching via blackbox differentiation of combinatorial solvers, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVIII 16, Springer, 2020, pp. 407–424.

[49] X. Liao, Y. Xu, H. Ling, Hypergraph neural networks for hypergraph matching, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 1266–1275.

[50] R. Wang, Z. Guo, S. Jiang, X. Yang, J. Yan, Deep learning of partial graph matching via differentiable top-k, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 6272–6281.

[51] K. Fu, S. Liu, X. Luo, M. Wang, Robust point cloud registration framework based on deep graph matching, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 8893–8902.

[52] K. Korotkov, J. Quintana, R. Campos, A. Jesús-Silva, P. Iglesias, S. Puig, J. Malvehy, R. Garcia, An improved skin lesion matching scheme in total body photography, IEEE journal of biomedical and health informatics 23 (2) (2018) 586–598.

[53] K. Korotkov, J. Quintana, S. Puig, J. Malvehy, R. Garcia, A new total body scanning system for automatic change detection in multiple pigmented skin lesions, IEEE transactions on medical imaging 34 (1) (2014) 317–338.

[54] M. Strakowska, M. Kociołek, Skin lesion matching algorithm for application in full body imaging systems, in: International Conference on Information Technologies in Biomedicine, Springer, 2022, pp. 222–233.

[55] M. H. Strzelecki, M. Strąkowska, M. Kozłowski, T. Urbańczyk, D. Wielowieyska-Szybińska, M. Kociołek, Skin lesion detection algorithms in whole body images, Sensors 21 (19) (2021) 6639.

[56] R. Marin, E. Corona, G. Pons-Moll, Nicp: Neural icp for 3d human registration at scale, in: European Conference on Computer Vision, 2024.

[57] P. J. Besl, N. D. McKay, Method for registration of 3-d shapes, in: Sensor fusion IV: control paradigms and data structures, Vol. 1611, Spie, 1992, pp. 586–606.

[58] F. Prada, M. Kazhdan, M. Chuang, A. Collet, H. Hoppe, Motion graphs for unstructured textured meshes, ACM Transactions on Graphics (TOG) 35 (4) (2016) 1–14.

[59] J. Munkres, Algorithms for the assignment and transportation problems, Journal of the society for industrial and applied mathematics 5 (1) (1957) 32–38.

[60] R. Wang, Z. Guo, W. Pan, J. Ma, Y. Zhang, N. Yang, Q. Liu, L. Wei, H. Zhang, C. Liu, Z. Jiang, X. Yang, J. Yan, Pygmtools: A python graph matching toolkit, Journal of Machine Learning Research 25 (33) (2024) 1–7.
URL `https://jmlr.org/papers/v25/23-0572.html`

[61] A. Saint, K. Cherenkova, G. Gusev, D. Aouada, B. Ottersten, et al., Bodyfitr: robust automatic 3d human body fitting, in: 2019 IEEE International Conference on Image Processing (ICIP), IEEE, 2019, pp. 484–488.

[62] V. Useini, S. Tanadini-Lang, Q. Lohmeyer, M. Meboldt, N. Andratschke, R. P. Braun, J. Barranco García, Automatized self-supervised learning for skin lesion screening, Scientific Reports 14 (1) (2024) 12697.

[63] Cloudcompare, gPL software (2023).
URL `http://www.cloudcompare.org/`

[64] B. D. Lucas, T. Kanade, An iterative image registration technique with an application to stereo vision, in: IJCAI'81: 7th international joint conference on Artificial intelligence, Vol. 2, 1981, pp. 674–679.

[65] C. Tomasi, T. Kanade, Detection and tracking of point, Int J Comput Vis 9 (137-154) (1991) 3.

[66] B. Di Veroli, R. Lederman, J. Sosna, L. Joskowicz, Graph-theoretic automatic lesion tracking and detection of patterns of lesion changes in longitudinal ct studies, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2023, pp. 106–115.