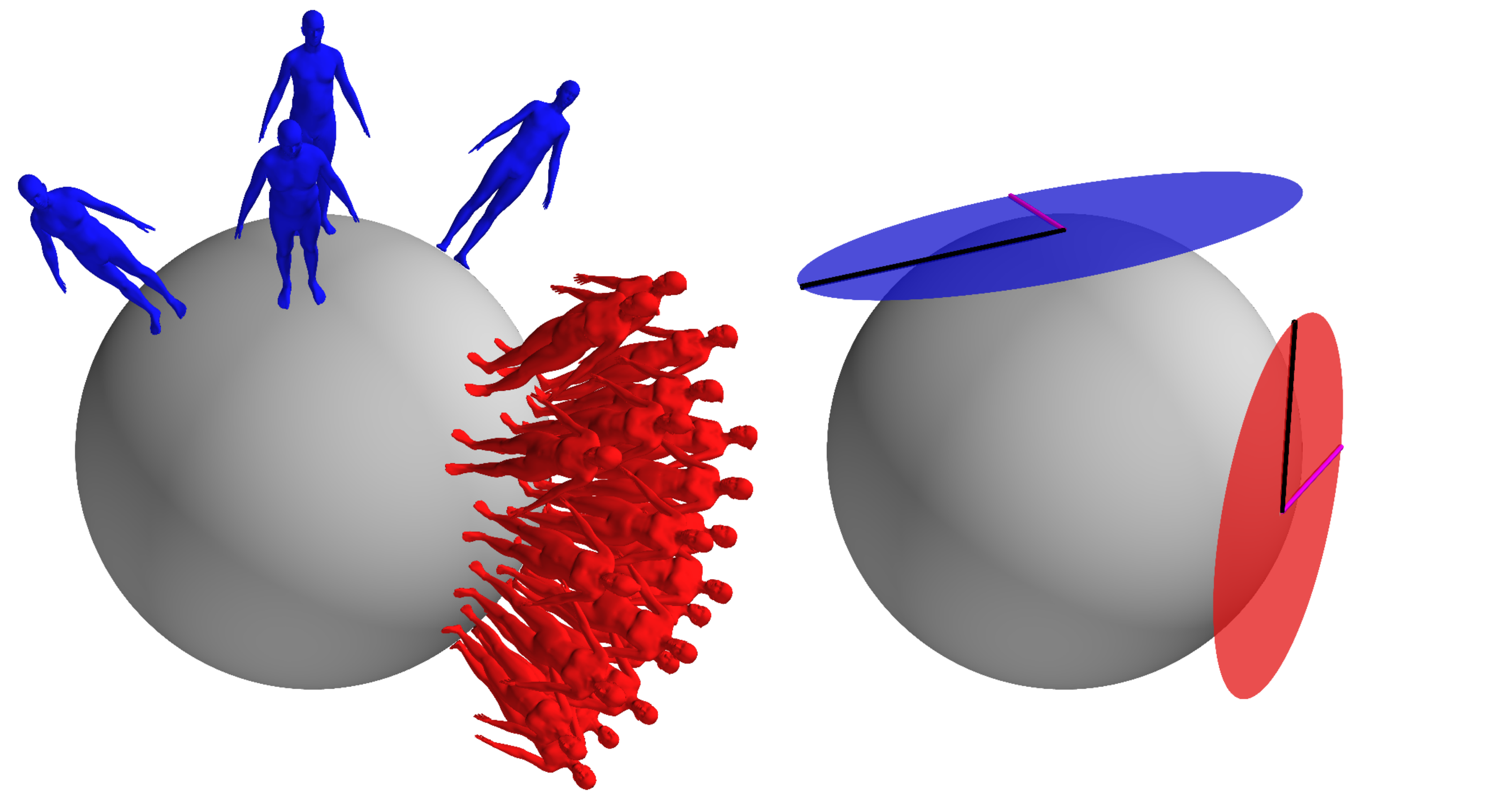
Model Transport: Towards Scalable Transfer Learning on Manifolds

Oren Freifeld — MIT

Søren Hauberg — Technical University of Denmark

Michael J. Black — MPI for Intelligent Systems

Introduction



- **Manifold-valued data** are ubiquitous in computer vision: surface normals, shape spaces, histogram-valued features, Symmetric Positive-Definite (SPD) matrices, the Grassmannian, etc.
- **Statistics on manifolds** is often done via tangent-space models; e.g., Gaussians, PCA, regression, classifiers, etc.
- One form of **transfer-learning** (TL) leverages a model learned in one region of $\mathbb{R}^n$ to improve a model in another region.
- **Goal:** Exploit TL ideas in modeling manifold-valued data.
- **Problem 1:** on a manifold, conventional $\mathbb{R}^n$-TL fails.
- Thought: Parallel Transport (PT) the data.
- **Problem 2:** this is not scalable.

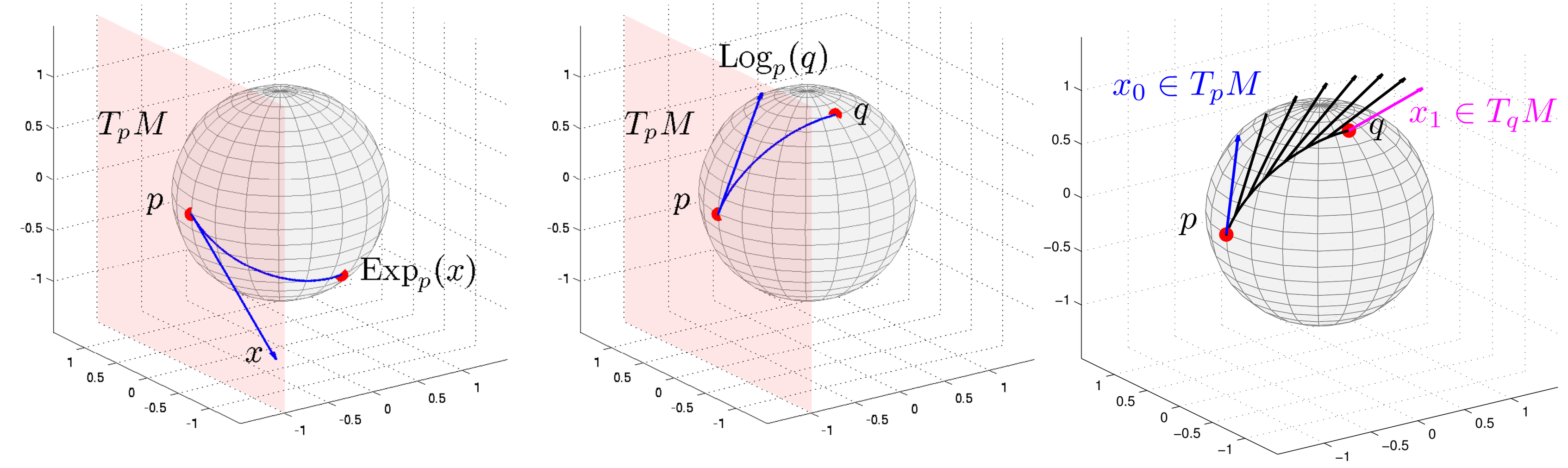
Solution: Transport the model, not the data

Key Contributions

- **Scalability:** We show how models can be transported. $\Longrightarrow$ # computations is fixed w.r.t. # data points; no need to store the data.
- **Optimality:** We show that for these models, PT and learning commute.

Parallel Transport (PT)

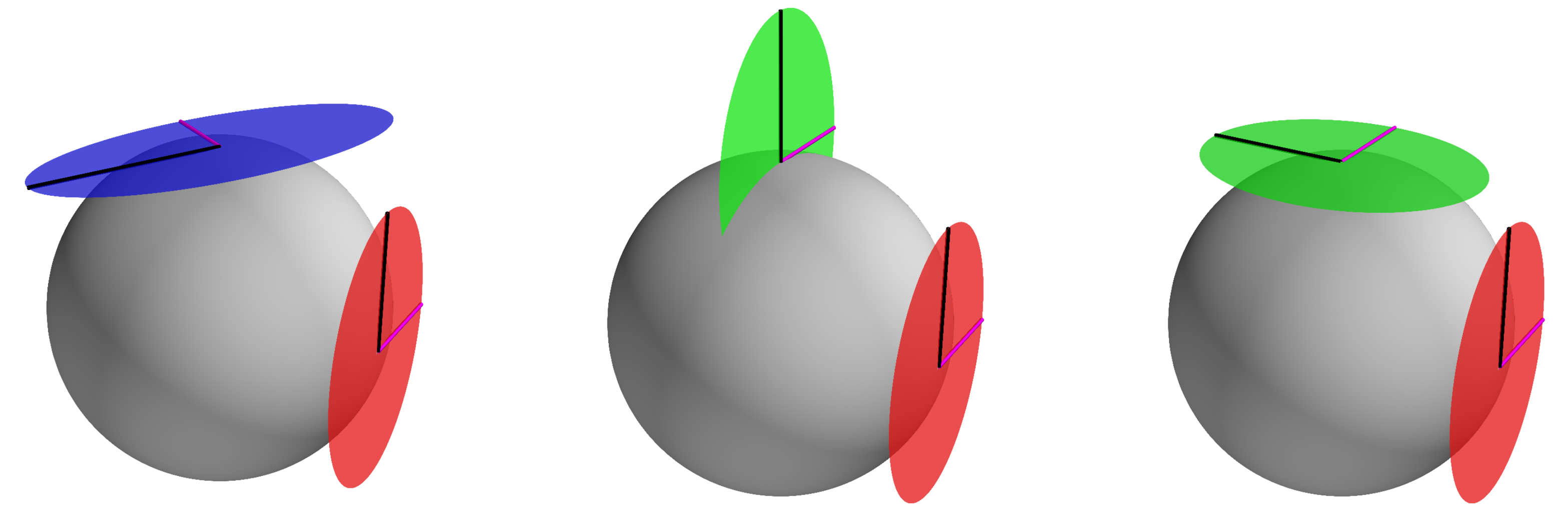
- An established tool to move vectors between tangent spaces.
- A *metric parallel transport* (MPT): inner-product-preserving PT.
- Both MPT and non-metric PT are **widely used in computer vision – but focus has been on expensive data transport**.



Acknowledgments

This work is supported in part by NIH-NINDS EUREKA (R01-NS066311). S.H. is supported in part by the Villum Foundation and the Danish Council for Independent Research (Natural Sciences).

Covariance/PCA Transport
(In short: keep the std. dev., transport the eig. vecs)

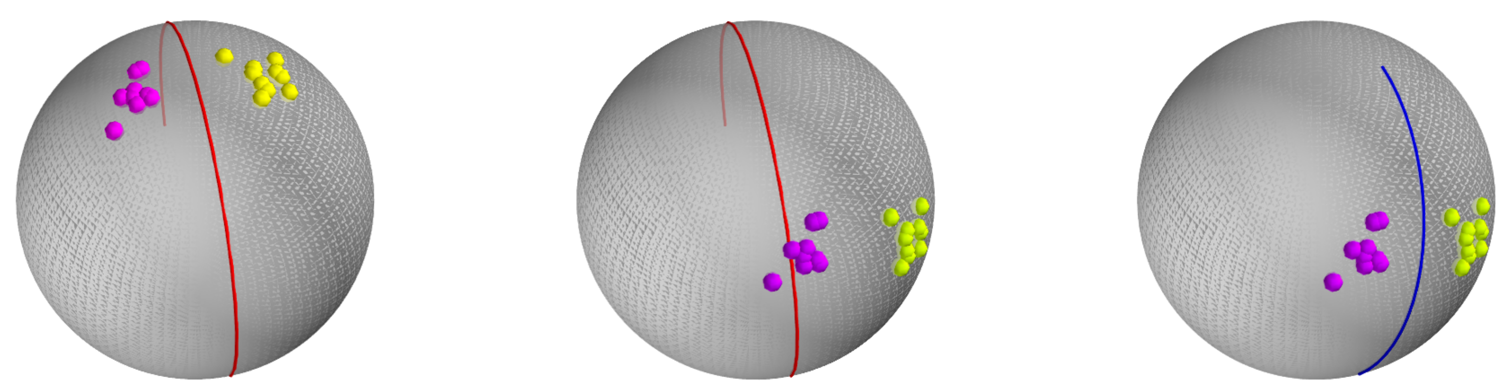


(a) Data models (b) $\mathbb{R}^n$-TL fails (c) Model transport

Figure: Model transport for improving covariance estimation.

- The transported model can improve the small-sample model.
- Point: only the eigen vectors need to be transported.

Regression/Classification Transport
(In short: transport the coefficient vector)



(a) Labeled training data and a logistic-regression classifier (b) Original classifier performs poorly in a another region (c) Classification transport

- Point: only a single vector needs to be transported.
- Similar results hold for **linear-regression** and **SVM** (PT the support vectors).

Why Does This Work?

- M: an n-dimensional manifold; $\tilde{x} \in T_qM$: the (metric) PT of $x \in T_pM$.

Data: $\{x_i\}_{i=1}^N \subset T_pM \quad X \triangleq [x_1, \dots, x_N] \qquad \tilde{X} \triangleq [\tilde{x}_1, \dots, \tilde{x}_N]$ (PT of the data)

Proposition 1 (Covariance/PCA Transport):

Let $VSU^T \overset{SVD}{=} X$, $VS^2V^T \overset{\text{eig. dec.}}{=} XX^T$, and $\tilde{V} \triangleq \overset{\text{PT of eigenvectors}}{[\tilde{v}_1, \dots, \tilde{v}_n]}$ where $[v_1, \dots, v_n] = V$. Then:

(a) $\tilde{V}SU^T \overset{SVD}{=} \tilde{X}$ and $\tilde{V}S^2\tilde{V}^T \overset{\text{eig. dec.}}{=} \tilde{X}\tilde{X}^T$.

(b) If $k < n$, then the k-dimensional PCA model of $\{\tilde{x}_i\}_{i=1}^N \subset T_qM$ is given by $\{\tilde{v}_i\}_{i=1}^k$ and $\{S_{i,i}/\sqrt{N-1}\}_{i=1}^k$.

- $\langle \cdot, \cdot \rangle_p : (x, y) \mapsto x^T A_p y$ and $\langle \cdot, \cdot \rangle_q : (x, y) \mapsto x^T A_q y$ are inner products on T_pM and T_qM ($A_p, A_q \in \mathrm{SPD}$). Data labels: $\{y_i\}_{i=1}^N \subset \mathbb{R}$. $L : T_pM \mapsto T_qM$: the linear map associated with an MPT. A linear regression model $T_pM \to \mathbb{R}$:

$$x \mapsto x^T\alpha + \alpha_0 = \langle x, A_p^{-1}\alpha \rangle_p + \alpha_0 \qquad \alpha_0 \in \mathbb{R} \qquad \alpha, A_p^{-1}\alpha \in T_pM\,.$$

- $l_i : \mathbb{R} \to \mathbb{R}_+$: a loss function associated with y_i; e.g., $l_i : \hat{y}_i \mapsto (\hat{y}_i - y_i)^2$.

Proposition 2 (Linear-Regression Transport):

$$\beta, \beta_0 = \underset{\alpha \in T_pM, \alpha_0 \in \mathbb{R}}{\arg\min} \sum_{i=1}^N l_i(x_i^T\alpha + \alpha_0) \Longrightarrow \gamma \triangleq A_q L A_p^{-1}\beta = \underset{\delta \in T_qM}{\arg\min} \sum_{i=1}^N l_i((Lx_i)^T\delta + \beta_0)$$

Application: Covariance Transport

Woman shapes improve a shape model of men while shapes of people with normal Body-Mass Index (BMI) improve a shape model of high-BMI people.

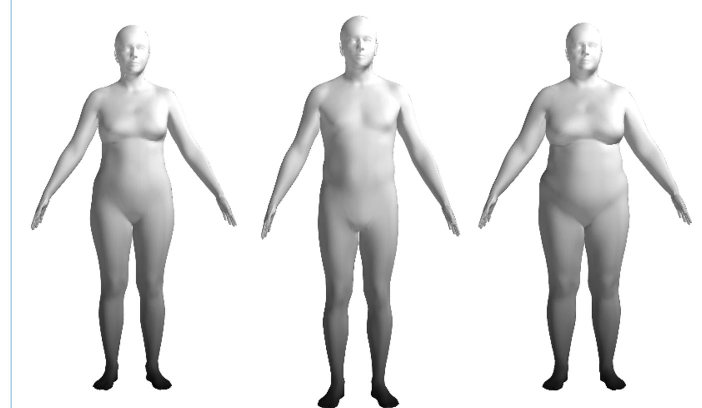


Figure: Mean shapes computed from: 1000 female shapes (left); 50 male shapes (middle); 50 shapes of women with high-BMI (right).

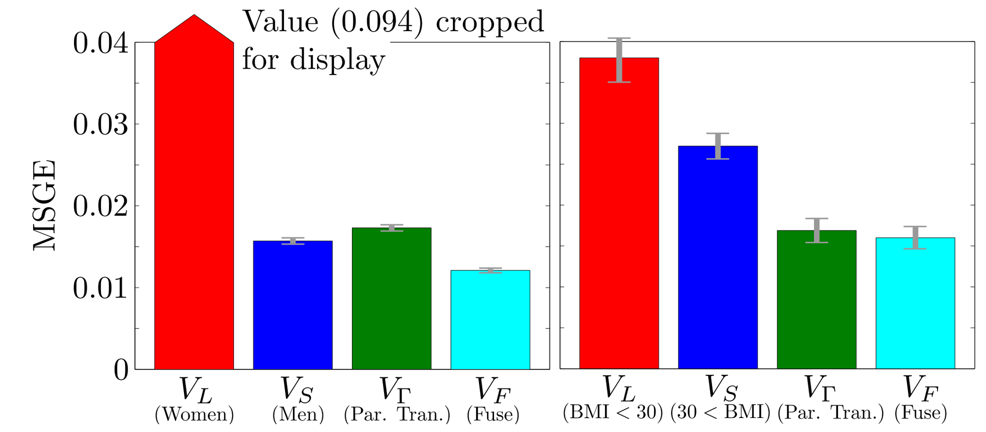
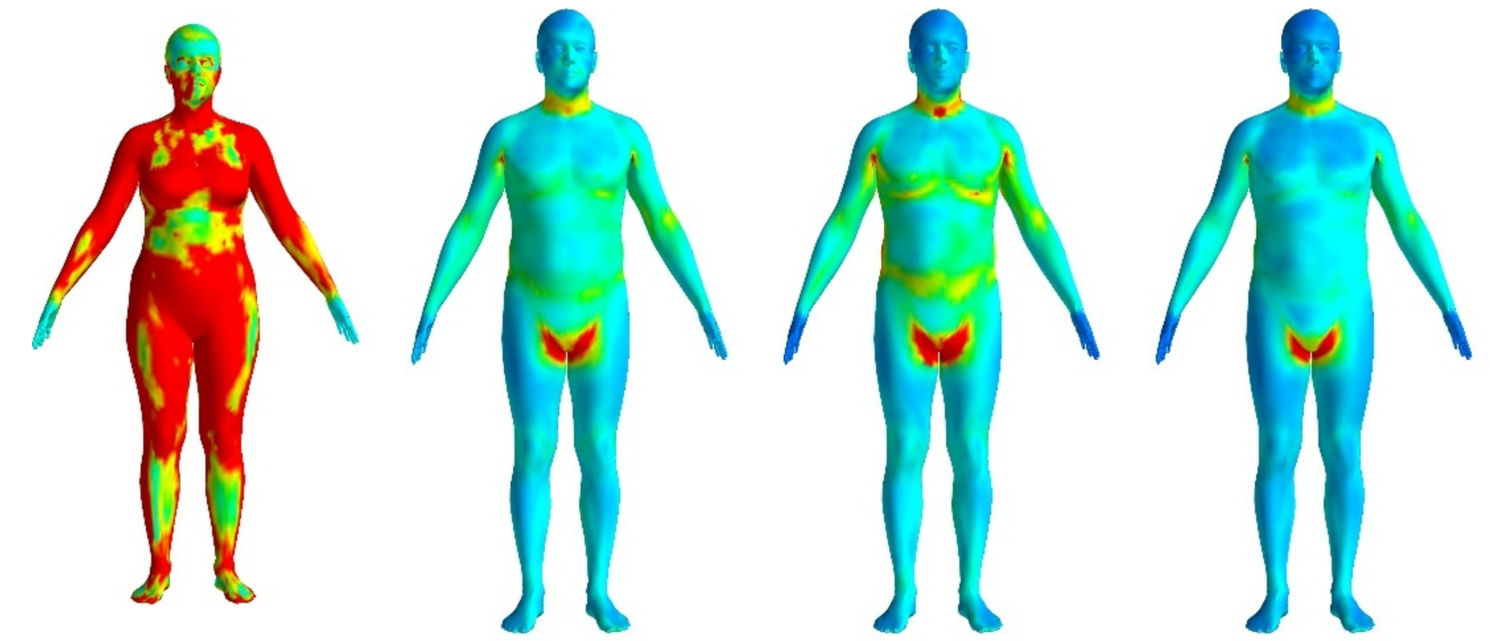


Figure: Reconstruction error measured in Squared Geodesic Error averaged over the test set as well as all of the 21550 triangles in the mesh. Geodesics here are w.r.t. the manifold from [Freifeld et al. ECCV '12].

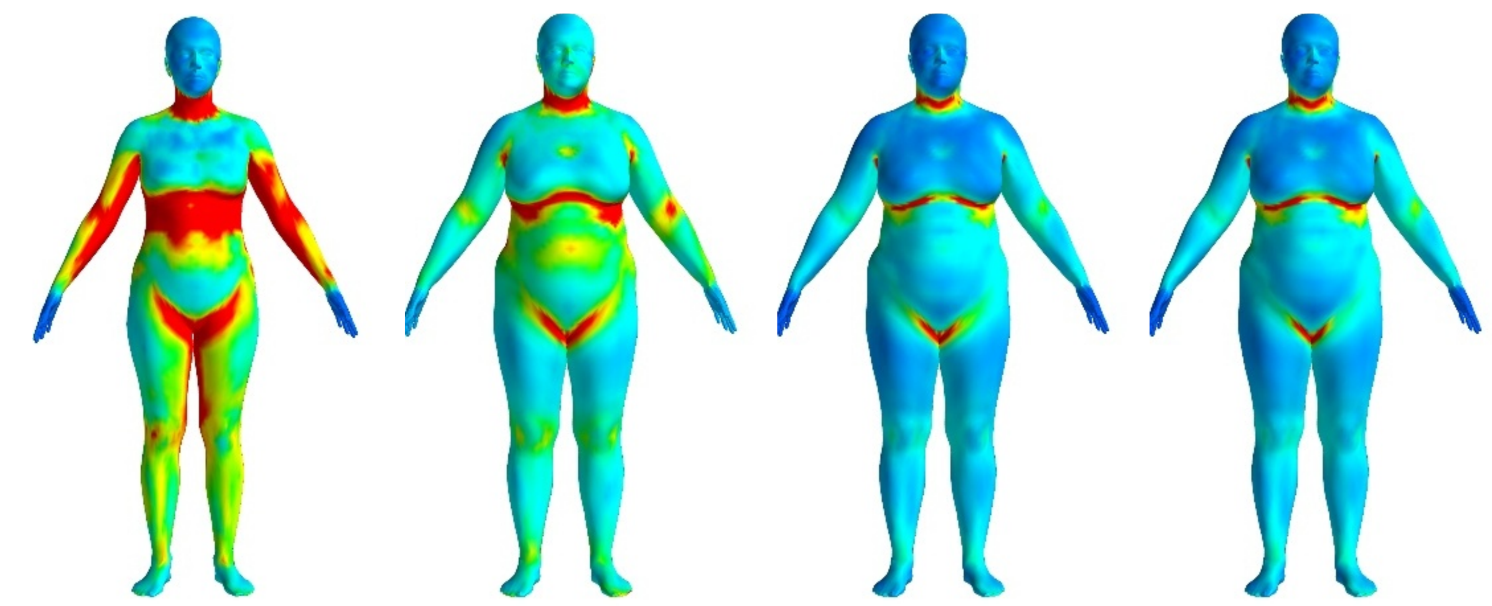
Improved Modeling of Male Shape using Female Shape



(a) V_L (women) (b) V_S (men) (c) V_Γ (PT) (d) V_F (Fusion)

Figure: Model mean error: Genders. Blue and red indicate small and large errors resp. Heat maps are overlaid over the points of tangency associated with the models: p for (a), and q for (b-e).

Improved Modeling of High-BMI Shape using Normal-BMI Shape



(a) V_L (Normal BMI) (b) V_S (High BMI) (c) V_Γ (PT) (d) V_F (Fusion)

Figure: Model mean error: BMI.

Application: Classifier Transport

Figure: Images from the 1st dataset. Left: class 1. Right: Class 2. Labels are known.

Figure: Images from the 2nd dataset. Left: class 1. Right: Class 2. Labels withheld.

- Features were encoded as SPD matrices. PT improves logistic-regression classifier results from 59% to 67%.
- Point: the same performance gain was obtained regardless whether we transported the data (168 vectors) or the model (a *single* vector).