MHEntropy: Entropy Meets Multiple Hypotheses for Pose & Shape Recovery

ICCV 2023

Glory Rongyu CHEN*

Linlin YANG*

Angela YAO

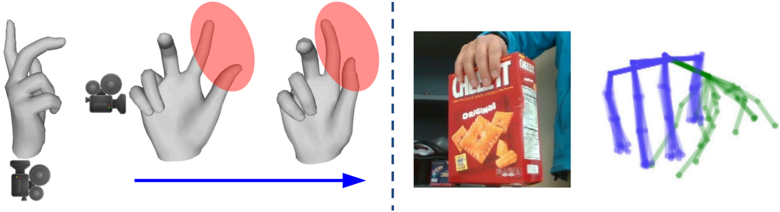
Sep, 2023

Computer Vision Machine Learning Group@NUS, SoC

Hints

- Why do we want Multi-Hypothesis?
- How does MH connect to Entropy?
- What are entropy effects & how to evaluate?

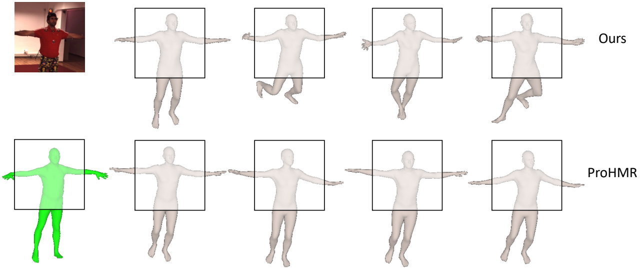
Common ambiguities



- Single-view RGB $\rightarrow$ human SMPL/hand MANO params
- Ambiguities: depth, occlusion/truncation & low image quality...
- Our goal: consistent w/ evidence – gold criterion: accurate poses on 2D visible while diverse & feasible on 3D occluded
- 1-v-*m* correspondence: deterministic ✗, probabilistic MH ✓

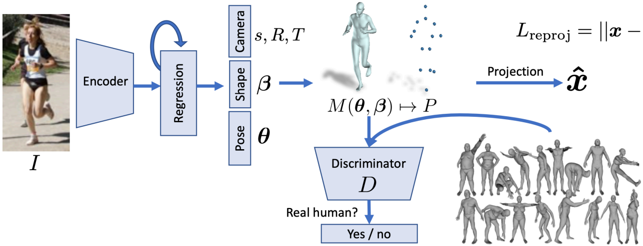
Lack of 1-v-m correspondence data

- Many methods [MDN, CVPR'19], [ProHMR, ICCV'21] trained w/ Maximum Likelihood Estimation require data, esp. 1-v-m DATA!
- 1-v-m paired data $(\mathbf{x}^{(i)}, \{\mathbf{y}^{(i,1)}, \mathbf{y}^{(i,2)}, \ldots\}) \ll$ uncond generation; usually only have one $(\mathbf{x}^{(i)}, \mathbf{y}^{(i,1)})$ since $\mathbf{x}^{(i)} \neq \mathbf{x}^{(j)}$
- Models implicitly learn 1-v-m across the dataset – mode collapse (not diverse) into 1-v-1 Det in a high dim [ProHMR, ICCV'21]
- Even worse under weak sup w/o $\mathbf{y}$, cannot apply MLE



What can be used except data? Knowledge!

- How the humans do? Guessing occluded while lifting into 3D



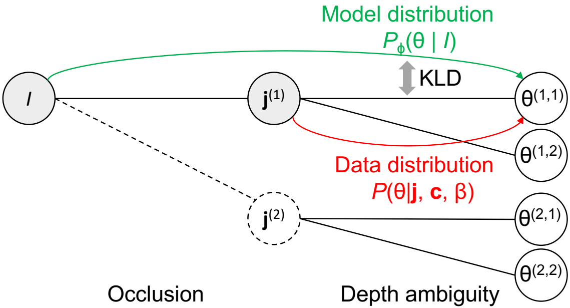
Follow human's feasibility priors: checking 2D proj & prior [HMR, CVPR'18]

- Many off-the-shelf general priors $P(\boldsymbol{\Theta})$ ✓

Diversity? Also knowledge about occluded

- Many possible positions w/o more cues – Maximum Entropy

- Our intuitions: use knowledge to define the target data distribution to alleviate reliance on annotations



A probabilistic framework: MHEntropy

- **Data distribution:** $P_{\mathrm{d}}(\boldsymbol{\Theta}|\mathbf{I})$ defined by 2D proj $P(\mathbf{j}|\boldsymbol{\Theta})$ (acc Laplace on vis + div Uniform on occ $\iff$ vis weighting) & prior $P(\boldsymbol{\Theta})$ knowledge, introduced by Bayes' rules,

$$p(\boldsymbol{\theta}|\mathrm{j}, \cancel{\mathrm{I}}) = p(\boldsymbol{\theta}|\mathrm{j}) \propto \delta(\mathrm{j}|\boldsymbol{\theta})p(\boldsymbol{\theta}),$$

$$p_{\mathrm{d}}(\boldsymbol{\theta}|\mathrm{I}) = \int_{\mathrm{j}} p(\boldsymbol{\theta}|\mathrm{j})p(\mathrm{j}|\mathrm{I})\mathrm{dj} \propto p(\boldsymbol{\theta})p_{\mathrm{proj}}(\mathrm{j} = \boldsymbol{\pi}|\mathrm{I})$$

- **Model distribution:** $P_{\boldsymbol{\phi}}(\boldsymbol{\Theta}|\mathbf{I})$ NFs
- **Distributional optimization:** $D_{\mathrm{KL}}(P_{\boldsymbol{\phi}}||P_{\mathrm{d}})$
- 3 terms: reconstruction, prior & missing entropy,

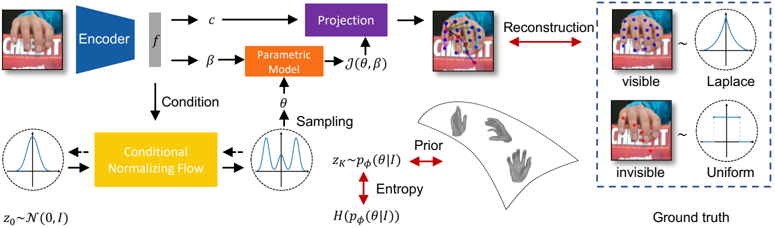
$$\text{rec: } \mathbb{E}_{p_{\boldsymbol{\phi}}(\boldsymbol{\theta}|\mathrm{I})}[\log p_{\mathrm{proj}}(\mathrm{j} = \boldsymbol{\pi}|\mathrm{I})],$$

$$\text{prior: } \mathbb{E}_{p_{\boldsymbol{\phi}}(\boldsymbol{\theta}|\mathrm{I})}[\log p(\boldsymbol{\theta})],$$

$$\text{ent: } -\mathbb{E}_{p_{\boldsymbol{\phi}}(\boldsymbol{\theta}|\mathrm{I})}[\log p_{\boldsymbol{\phi}}(\boldsymbol{\theta}|\mathrm{I})]$$

- The essential entropy encourages meaningful diversity when it can, to further optimize the objective

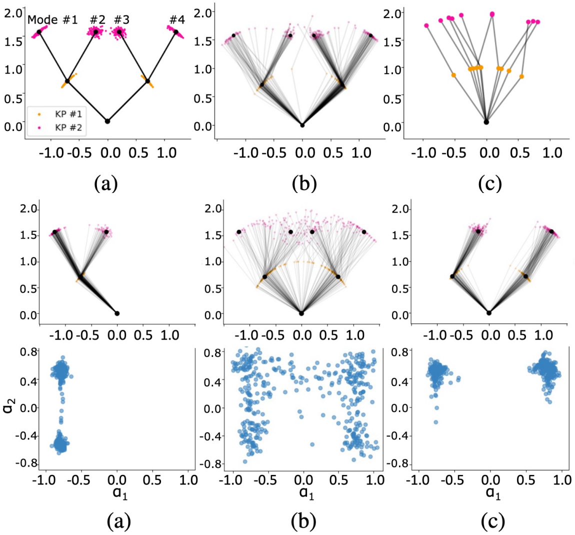
Framework



- Extract features **f**
- Sample $\boldsymbol{\theta}$ from NFs conditional on **f**
- Compute 3 losses & BackPropagate

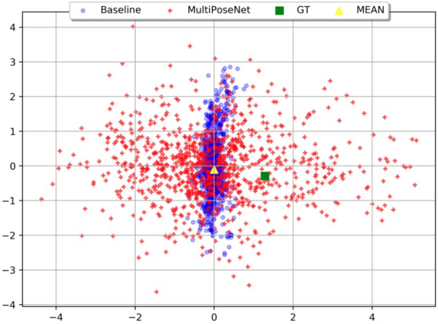
From our proposed probabilistic framework,

- [ProHMR, ICCV'21] no entropy, biased distributions
- [WS3DPG, BMVC'20] GANs + heuristic regularizations, hard to optim
- [CMVAE, CVPR'18] less expressive unimodal distributions
- [Others] 'data-driven'



3 terms

A comprehensive MH evaluation except BH



[CVAE, ICCV'19]

- The Best Hypothesis w.r.t. 1 possible annot, min MPJPE
- Gold criteria

  [Acc] All Hypothesis on vis

  [Div] Per Joint Diversity (STD) on 2D/3D vis/occ parts

  [Div|Acc] Relative Diversity, the most certain / uncertain $\frac{\text{PJD}_{\text{2d vis}}}{\text{PJD}_{\text{3d occ}}}$

(A-)H36M acc & div results

Supervision		MH	H36M	AH36M
2D Vis	HMR		67.4	85.2
	ProHMR	✓	64.3	82.6
	Ours	✓	**51.3**	**66.4**
3D	HMR		56.8	-
	SPIN		41.1	-
	MDN	✓	42.7	69.5
	CVAE	✓	46.2	75.1
	Multi-bodies	✓	42.2	64.2
	ProHMR	✓	**36.8**	<u>60.1</u>
	Ours	✓	**36.8**	**50.6**

Table 2. PA-MPJPE (mm) of BH results on H36M and its ambiguous version AH36M under the supervision of visible 2D keypoints (2D Vis) and 3D keypoints (3D) with $n = 25$.

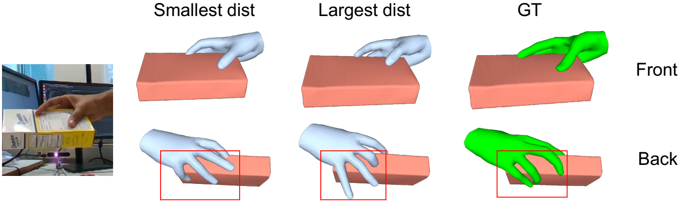
Supervision		AH (pix)↓	PJD		RD↓
			2D Vis	3D Occ	
2D Vis	ProHMR	10.92	0.06	0.26	0.23
	Ours	**9.75**	4.56	64.05	**0.07**
3D	ProHMR	13.38	3.98	24.27	0.16
	Ours	**10.73**	4.23	47.95	**0.09**

Table 3. Diversity metrics on AH36M under the supervision of visible 2D keypoints (2D Vis) and 3D keypoints (3D).

Meet gold criteria more under both weak/strong sup

Where our MH can be used?

- Post-selection, *e.g.*, Hand-Object Interaction grasp & multi-view



- Fitting evidence as a better prior
- Additional consistency reconstruction w/ more knowledge (*e.g.*, masks)
- Consider more ambiguities like image blur

Takeaways

- A probabilistic framework for a partial weak sup setting makes use of knowledge & derives a missing entropy
- The flexible framework can incorporate more consistency & ambiguities for meaningful diversity only on uncertain
- Comprehensive MH evaluation

Acknowledgement

- CVPR’23 & ICCV’23 reviewers
- Dr. Chen LI (NUS) working in MH
- Junpeng HU (SJTU) for math verification
- (CVML) Ziwei YU, Tianyu GAO, Dr. Guodong DING, & Ha Linh, *etc.*

References

[MDN, CVPR'19] C. Li and G. H. Lee, "Generating multiple hypotheses for 3D human pose estimation with mixture density network," in CVPR, 2019.

[ProHMR, ICCV'21] N. Kolotouros, G. Pavlakos, D. Jayaraman, and K. Daniilidis, "Probabilistic modeling for human mesh recovery," in ICCV, 2021.

[HMR, CVPR'18] A. Kanazawa, M. J. Black, D. W. Jacobs, and J. Malik, "End-to-end recovery of human shape and pose," in CVPR, 2018.

[WS3DPG, BMVC'20] C. Li and G. H. Lee, "Weakly supervised generative network for multiple 3D human pose hypotheses," in BMVC, 2020.

[CMVAE, CVPR'18] A. Spurr, J. Song, S. Park, and O. Hilliges, "Cross-modal deep variational hand pose estimation," in CVPR, 2018.

[CVAE, ICCV'19] S. Sharma, P. T. Varigonda, P. Bindal, A. Sharma, and A. Jain, "Monocular 3D human pose estimation by generation and ordinal ranking," in ICCV, 2019.

Thank You for Attention!