Photo Wake-Up:

3D Character Animation from a Single Photo

ECCV 2018

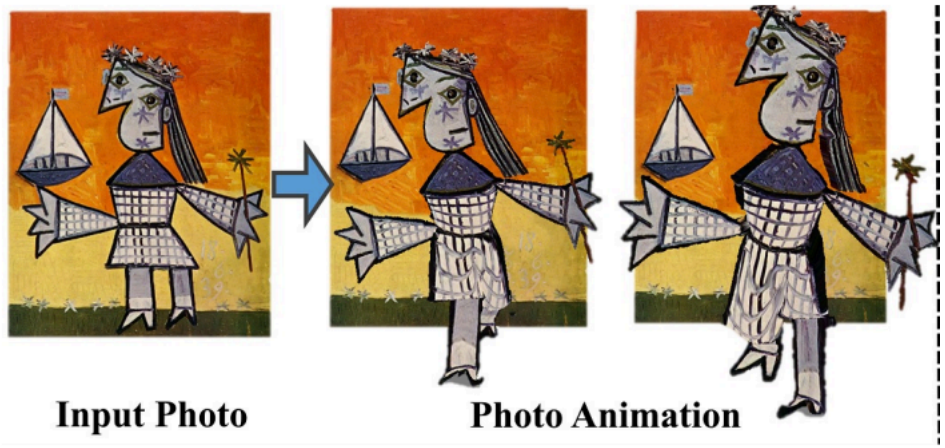
00. INTRO

Photo Wake-up

https://youtu.be/G63goXc5MyU

00. INTRO

Photo Wake-up



Input Photo Photo Animation

Augmented Reality

00. INTRO

Photo Wake-up

an application of viewing and animating humans **in single photos** in 3D

a novel 2D warping method to deform a posable template body model
to fit the person's complex silhouette to create an animatable mesh

a method for handling partial **self occlusions**

00. INTRO

Photo Wake-up

1. Mesh construction

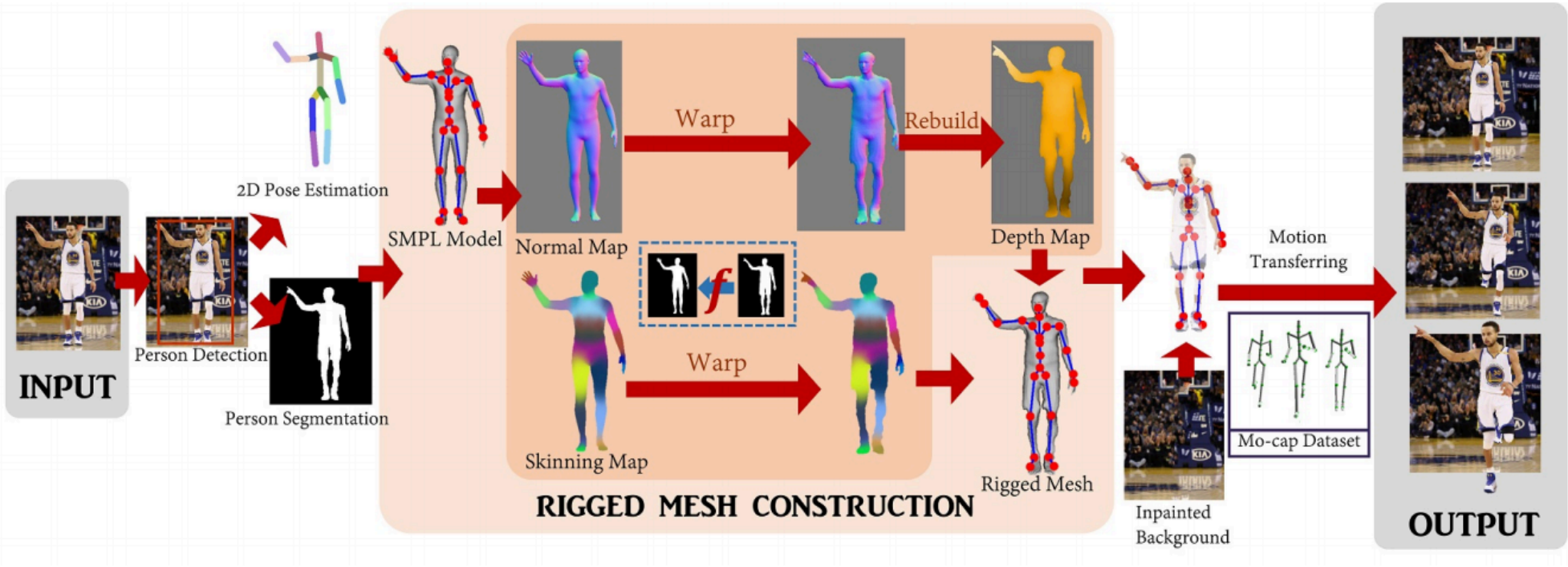
2. Self occlusion

3. Final steps

1. Mesh construction & Rigging

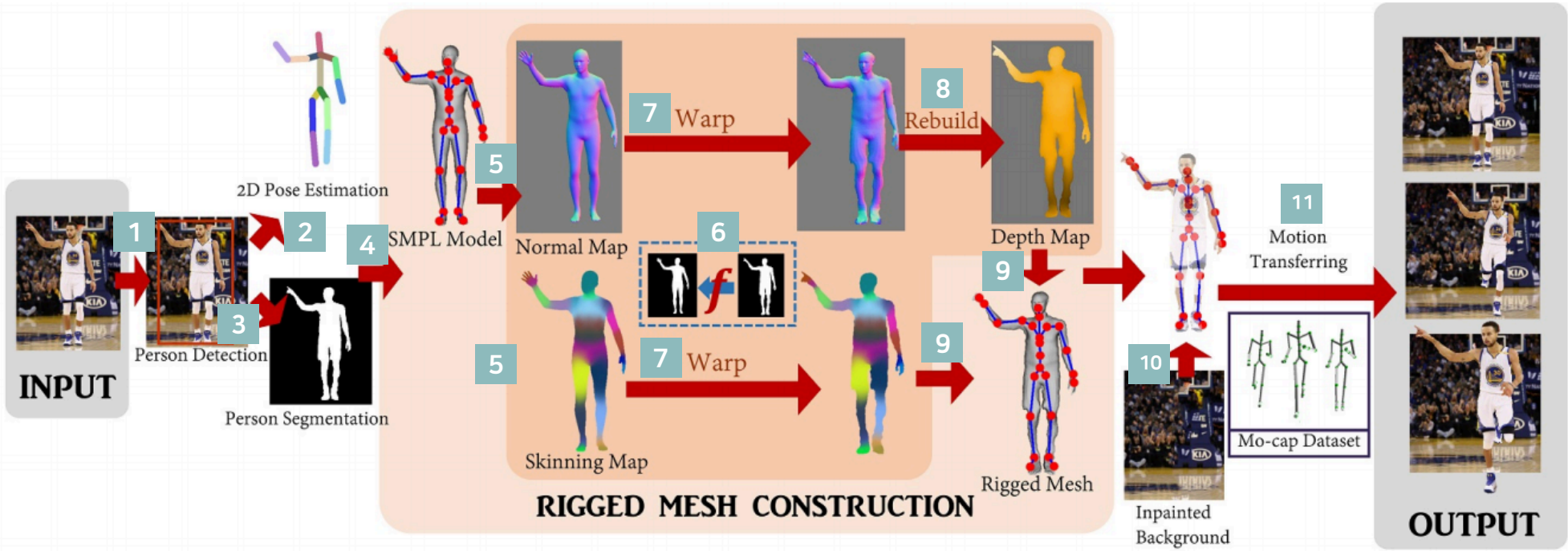
01. Mesh construction & Rigging

Introduction



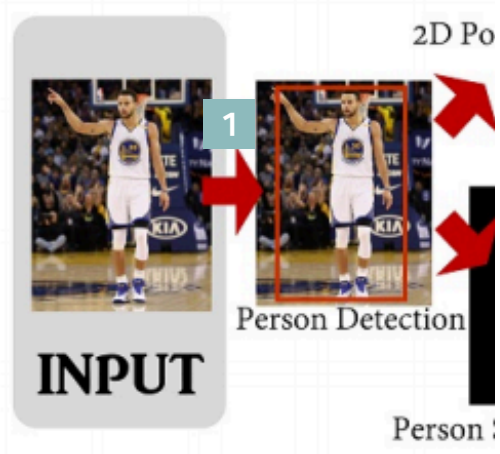
01. Mesh construction & Rigging

Introduction



01. Mesh construction & Rigging

1) Person detection



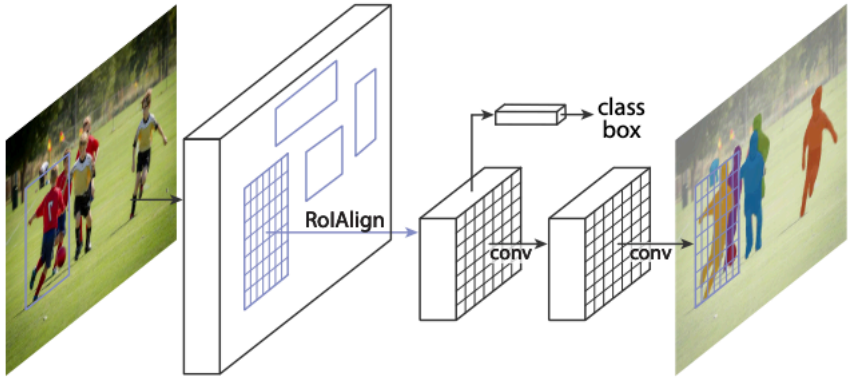


Figure 1. The **Mask R-CNN** framework for instance segmentation.

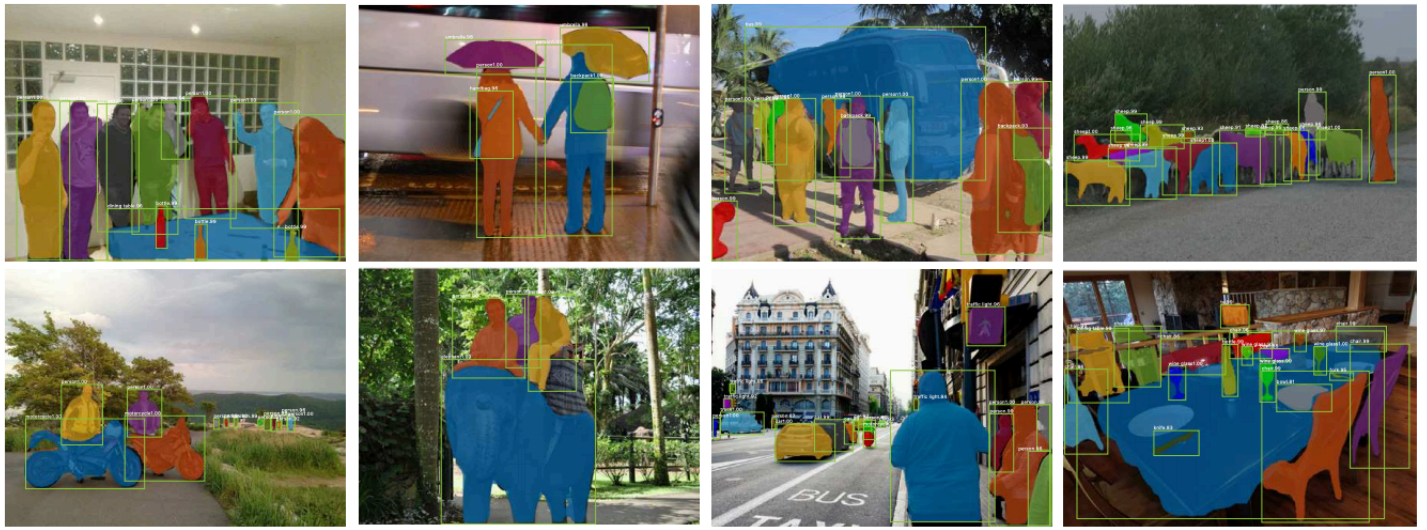
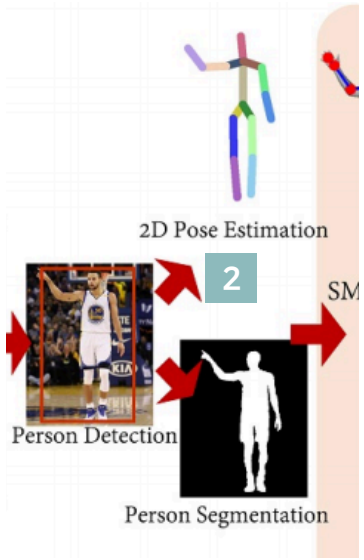


Figure 2. **Mask R-CNN** results on the COCO test set. These results are based on ResNet-101 [15], achieving a *mask* AP of 35.7 and running at 5 fps. Masks are shown in color, and bounding box, category, and confidences are also shown.

[25] He, Kaiming, et al. "Mask r-cnn." IEEE. 2017.

01. Mesh construction & Rigging

2) 2D Pose estimation



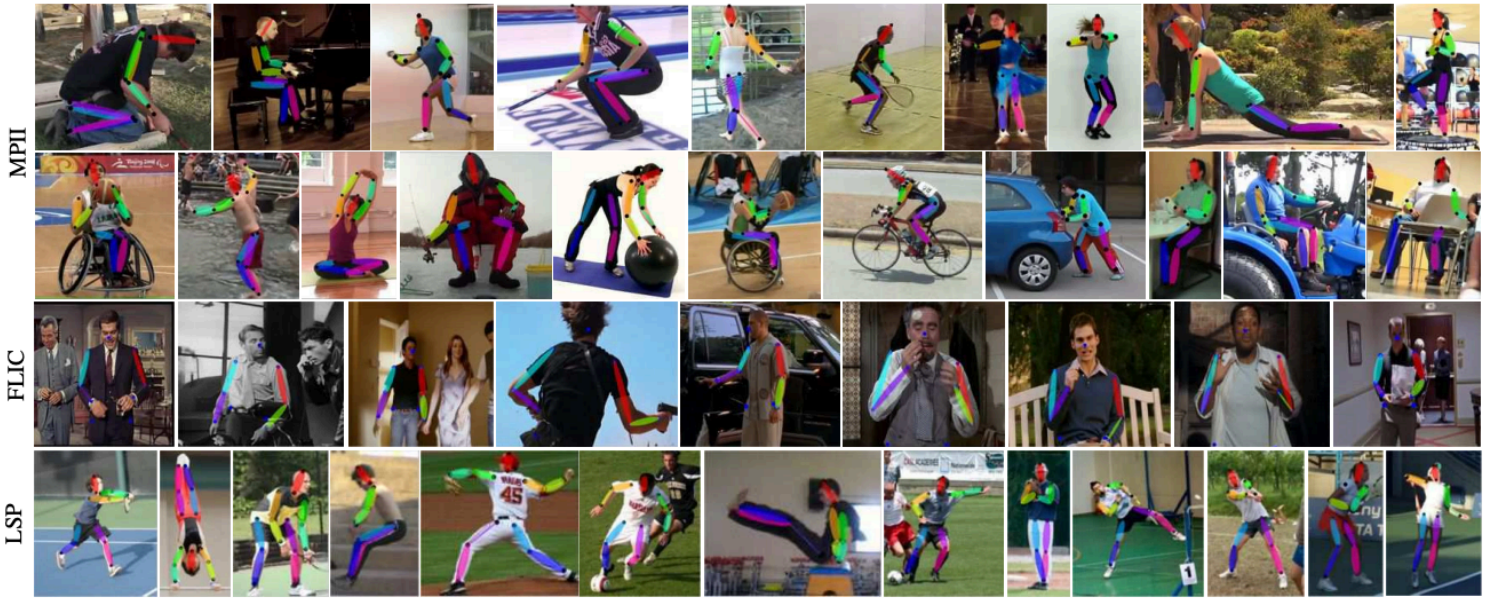
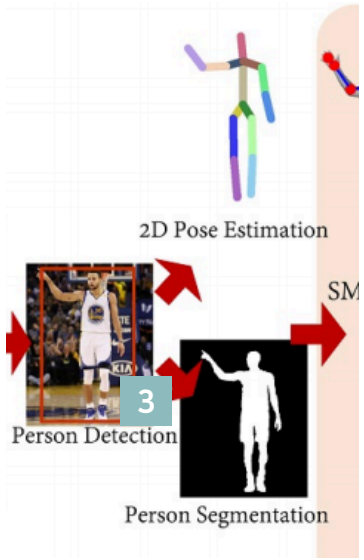


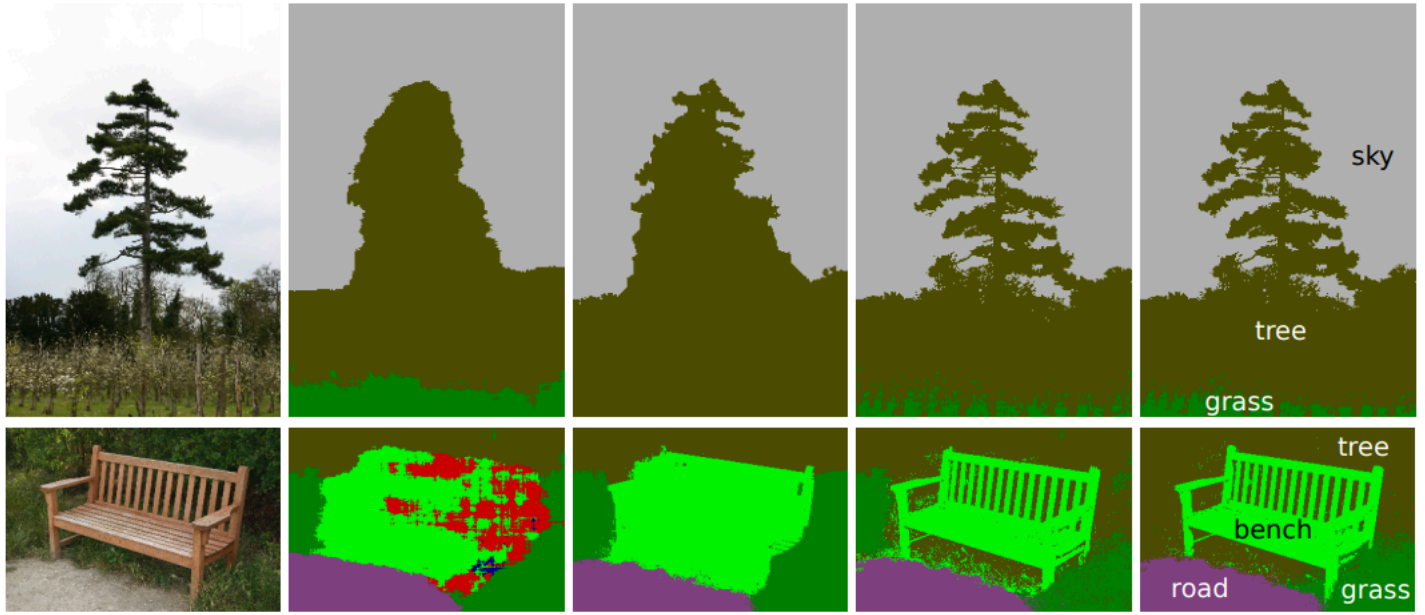
Figure 10: Qualitative results of our method on the MPII, LSP and FLIC datasets respectively. We see that the method is able to handle non-standard poses and resolve ambiguities between symmetric parts for a variety of different relative camera views.

[50] Wei, Shih-En, et al. "Convolutional pose machines." IEEE. 2016.

01. Mesh construction & Rigging

3) Person Segmentation





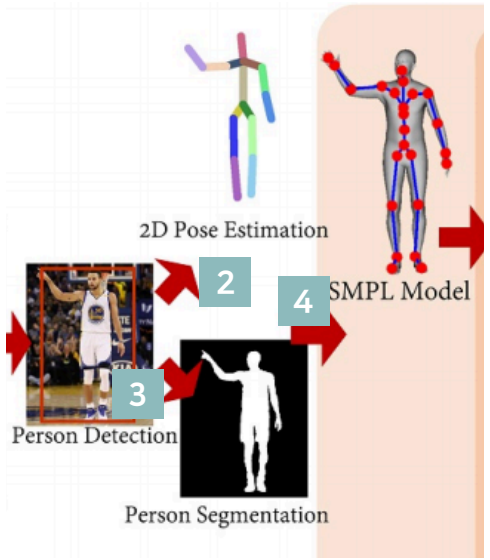
(a) Image (b) Unary classifiers (c) Robust P^n CRF (d) Fully connected CRF, MCMC inference, 36 hrs (e) Fully connected CRF, our approach, 0.2 seconds

Figure 1: Pixel-level classification with a fully connected CRF. (a) Input image from the MSRC-21 dataset. (b) The response of unary classifiers used by our models. (c) Classification produced by the Robust P^n CRF [9]. (d) Classification produced by MCMC inference [17] in a fully connected pixel-level CRF model; the algorithm was run for 36 hours and only partially converged for the bottom image. (e) Classification produced by our inference algorithm in the fully connected model in 0.2 seconds.

[35] Krähenbühl, Philipp, and Vladlen Koltun. "Efficient inference in fully connected crfs with gaussian edge potentials." Advances in neural information processing systems. 2011.

01. Mesh construction & Rigging

4) SMPL model



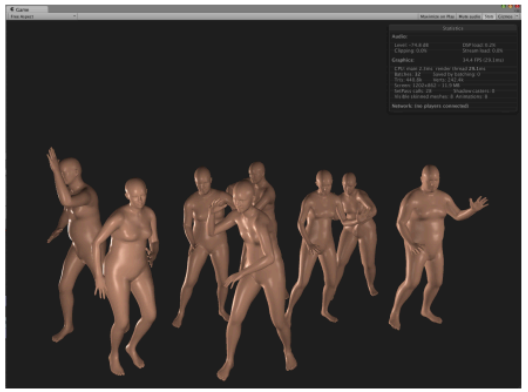
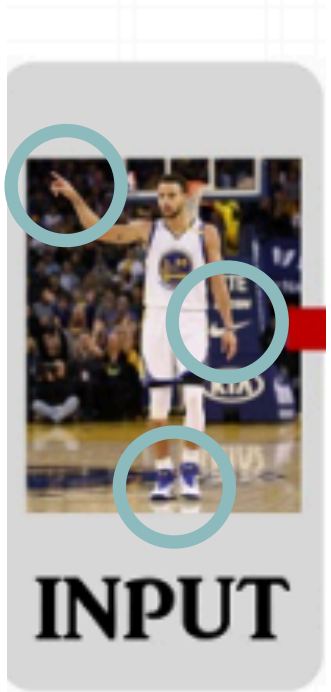


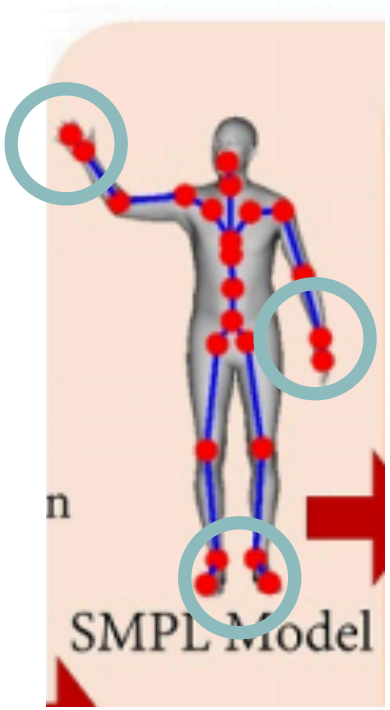
Figure 1: *SMPL is a realistic learned model of human body shape and pose that is compatible with existing rendering engines, allows animator control, and is available for research purposes. (left) SMPL model (orange) fit to ground truth 3D meshes (gray). (right) Unity 5.0 game engine screenshot showing bodies from the CAESAR dataset animated in real time.*

[36] dsd MLA Loper, Matthew, et al. "SMPL: A skinned multi-person linear model." ACM transactions on graphics (TOG) 34.6 (2015): 248.

01. Mesh construction & Rigging

4) SMPL model - 모델의 한계





[36] dsd MLA Loper, Matthew, et al. "SMPL: A skinned multi-person linear model." ACM transactions on graphics (TOG) 34.6 (2015): 248.

01. Mesh construction & Rigging

4) SMPL model - Solution

Bogo, Alldieck : SMPL모델을 만든 다음 그 모델이 실루엣에 맞도록 세부 조정

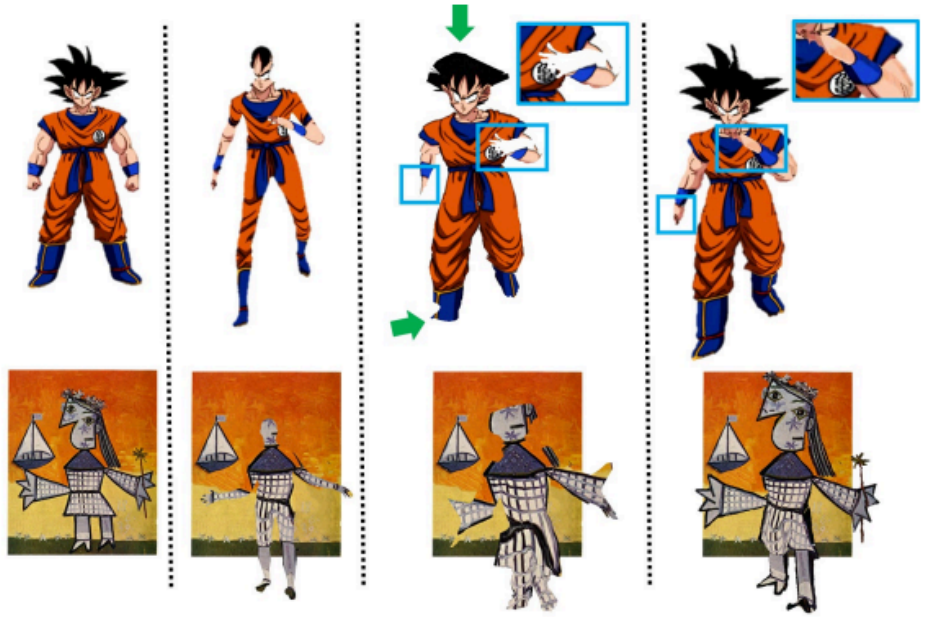
2D approach

- Match the person shilhouette in the original image
- Warp the projected SMPL normal & skinning map
- Construct front & back view and lift them into 3D along with 3D skeleton

[6] Bogo, Federica, et al. "Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image." European Conference on Computer Vision. Springer, Cham, 2016.
[10] Alldieck, Thiemo, et al. "Detailed human avatars from monocular video." IEEE, 2018.

01. Mesh construction & Rigging

4) SMPL model - solution



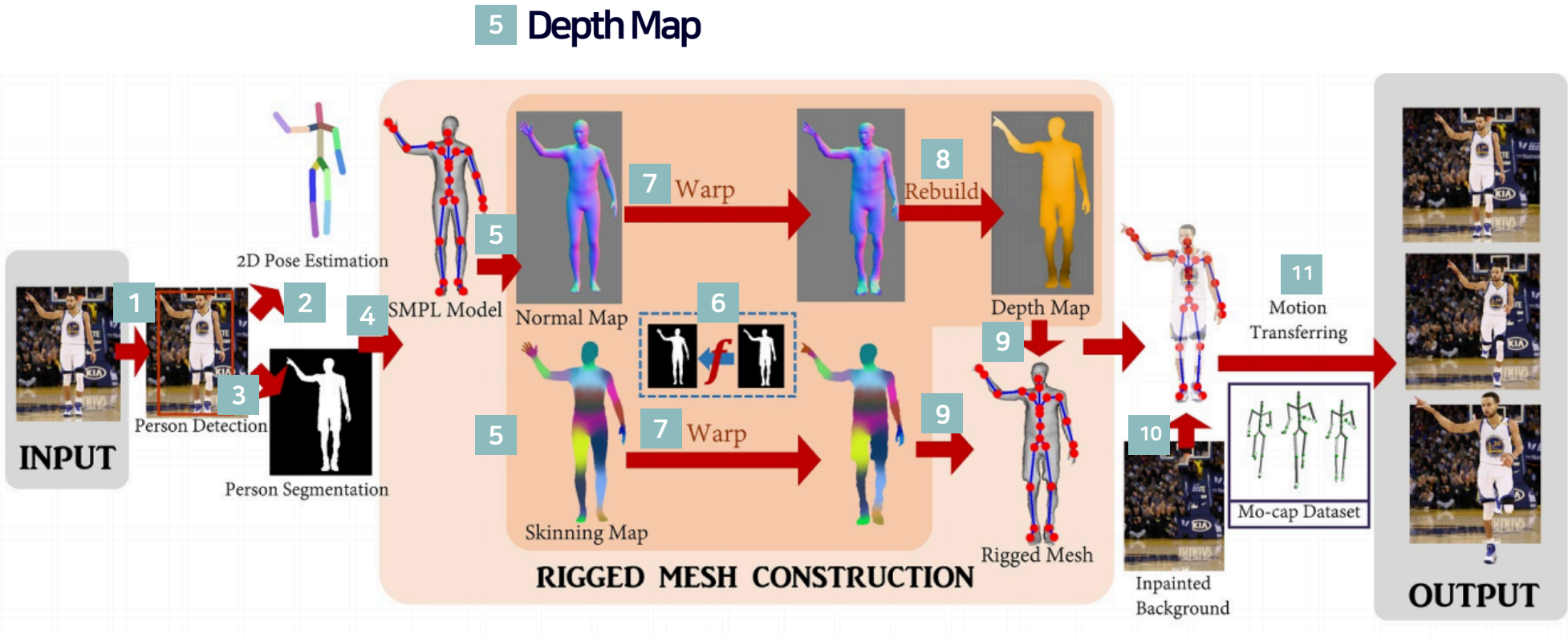
(a) input (b) Bogo et al. (c) Alldieck et al. (d) our method

Figure 11: Comparison with [16, 11]. (a) Input photo. (b) A fitted SMPL model [16]. (c) A deformed mesh using [11]. The mesh fails to deform hair and shoes (green arrows) and fingers (blue box). (d) Our mesh. *Photo credits: [2, 8]*

[6] Bogo, Federica, et al. "Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image." European Conference on Computer Vision. Springer, Cham, 2016.
[10] Alldieck, Thiemo, et al. "Detailed human avatars from monocular video." IEEE, 2018.

01. Mesh construction & Rigging

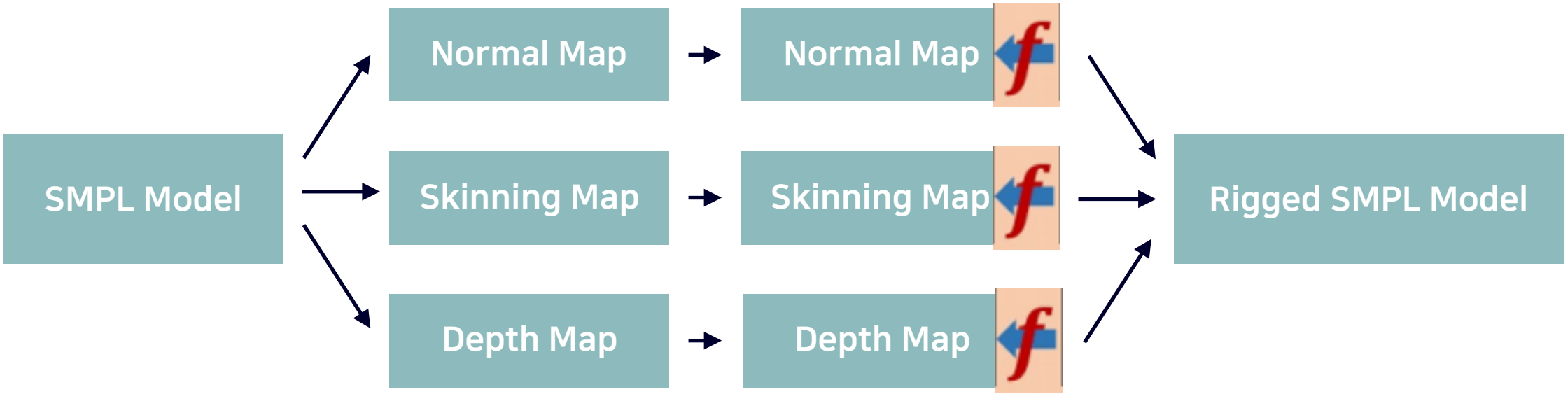
5) Normal map, Skinning map, Depth map



[6] Bogo, Federica, et al. "Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image." European Conference on Computer Vision. Springer, Cham, 2016.
[10] Alldieck, Thiemo, et al. "Detailed human avatars from monocular video." IEEE, 2018.

01. Mesh construction & Rigging

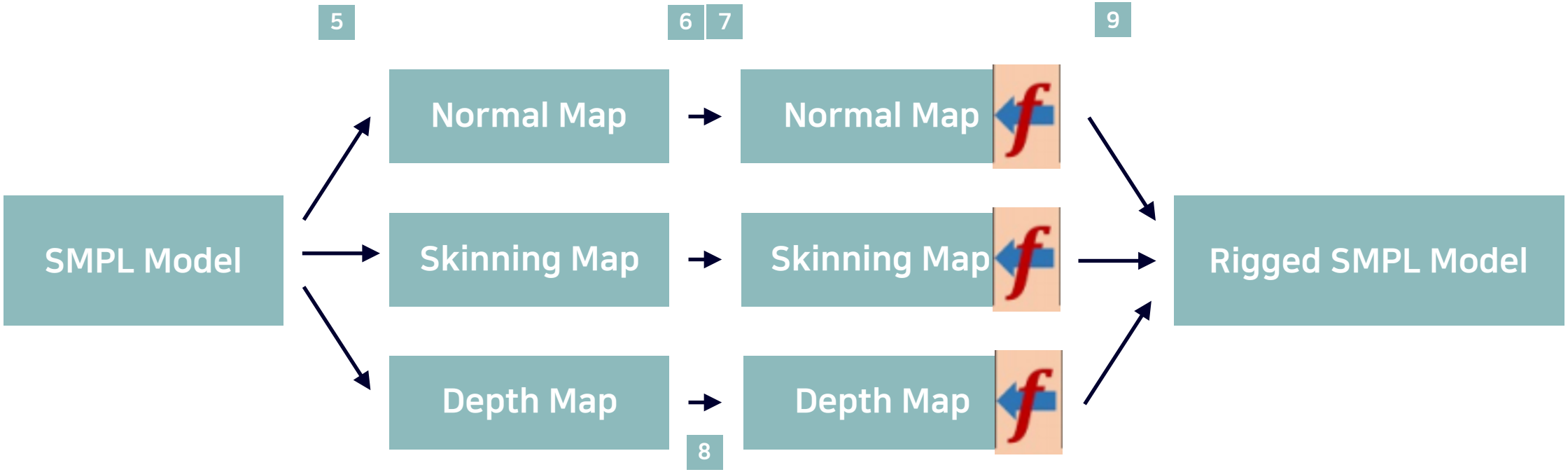
5) Normal map, Skinning map, Depth map



[6] Bogo, Federica, et al. "Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image." European Conference on Computer Vision. Springer, Cham, 2016.
[10] Alldieck, Thiemo, et al. "Detailed human avatars from monocular video." IEEE, 2018.

01. Mesh construction & Rigging

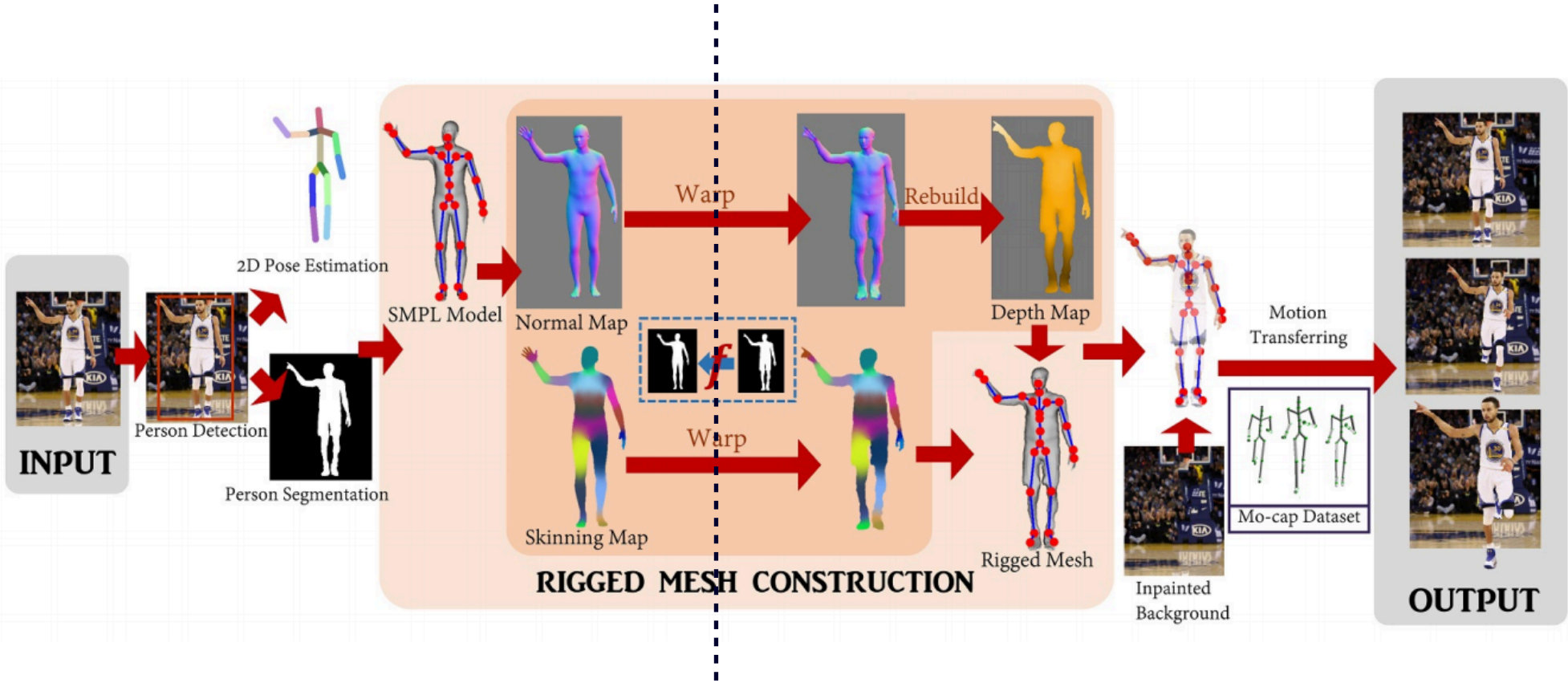
5) Normal map, Skinning map, Depth map



[6] Bogo, Federica, et al. "Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image." European Conference on Computer Vision. Springer, Cham, 2016.
[10] Alldieck, Thiemo, et al. "Detailed human avatars from monocular video." IEEE, 2018.

01. Mesh construction & Rigging

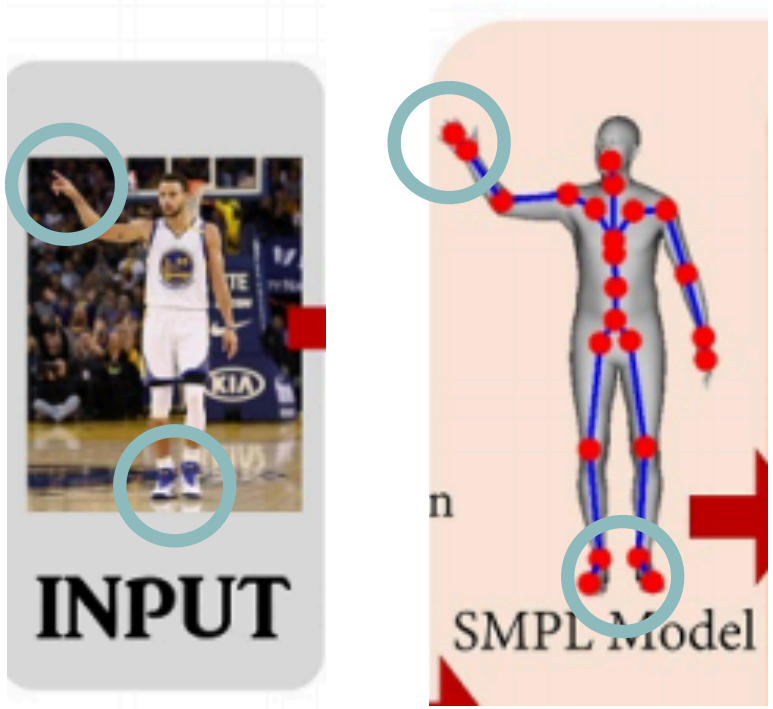
5) Normal map, Skinning map, Depth map



[6] Bogo, Federica, et al. "Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image." European Conference on Computer Vision. Springer, Cham, 2016.
[10] Alldieck, Thiemo, et al. "Detailed human avatars from monocular video." IEEE, 2018.

01. Mesh construction & Rigging

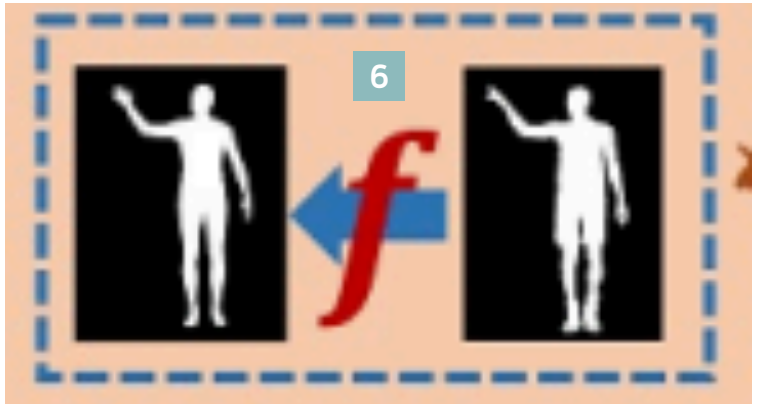
6) Inverse Warp



(a) (b)

01. Mesh construction & Rigging

6) Inverse Warp



(a) (b)

$$x = \sum_{i=0}^{m-1} \lambda_i(x) p_i \tag{6}$$

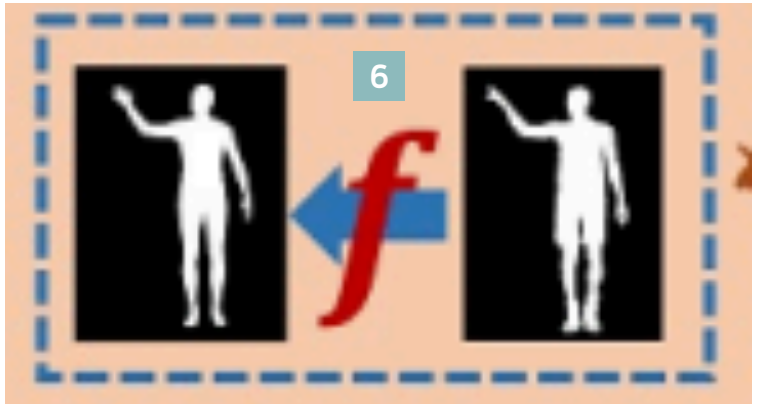
$$p_i \rightarrow p_{\phi[i]}^{\mathrm{SMPL}}. \tag{7}$$

$$f(x) = \sum_{i=0}^{m-1} \lambda_i(x) p_{\phi[i]}^{\mathrm{SMPL}}. \tag{8}$$

[23] Floater, Michael S. "Mean value coordinates." Computer aided geometric design 20.1 (2003): 19-27.

01. Mesh construction & Rigging

7) Boundary matching



(a) (b)

$$\underset{\phi[0],\ldots,\phi[m-1]}{\arg\min} \sum_{i=0}^{m-1} D(p_i, p_{\phi[i]}^{\mathrm{SMPL}}) + T(\phi[i], \phi([i+1])) \tag{9}$$

where

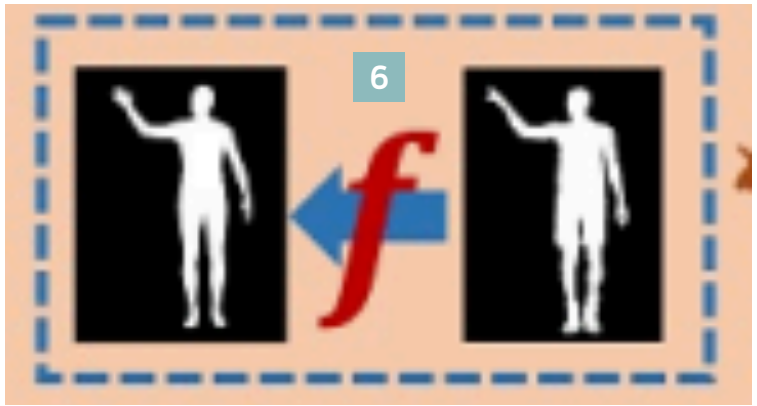
$$D(p_i, p_{\phi[i]}^{\mathrm{SMPL}}) = \|p_i - p_{\phi[i]}^{\mathrm{SMPL}}\|_2 \tag{10}$$

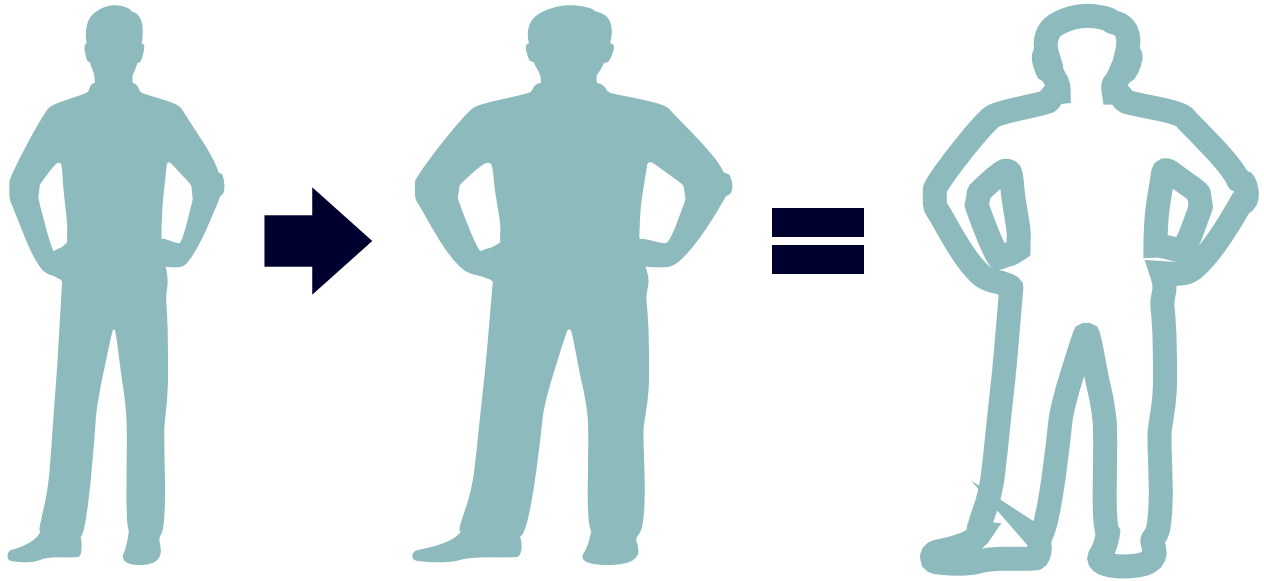
and

$$T(\phi[i], \phi[i+1]) = \begin{cases} 1, & \text{if } 0 \le \phi[i+1] - \phi[i] \le \kappa \\ \infty, & \text{otherwise} \end{cases} \tag{11}$$

01. Mesh construction & Rigging

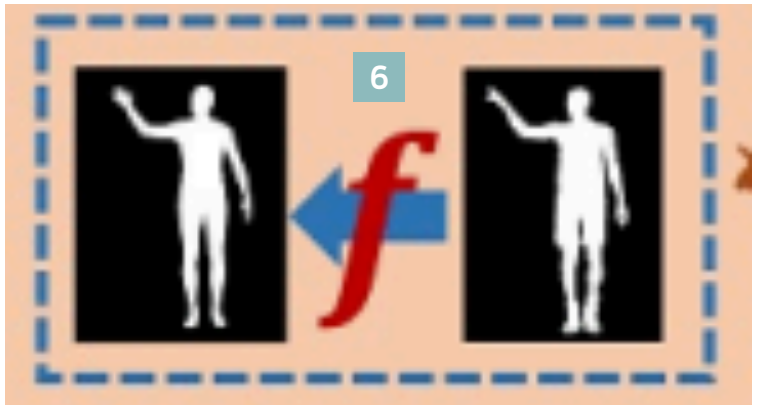
7) Boundary matching



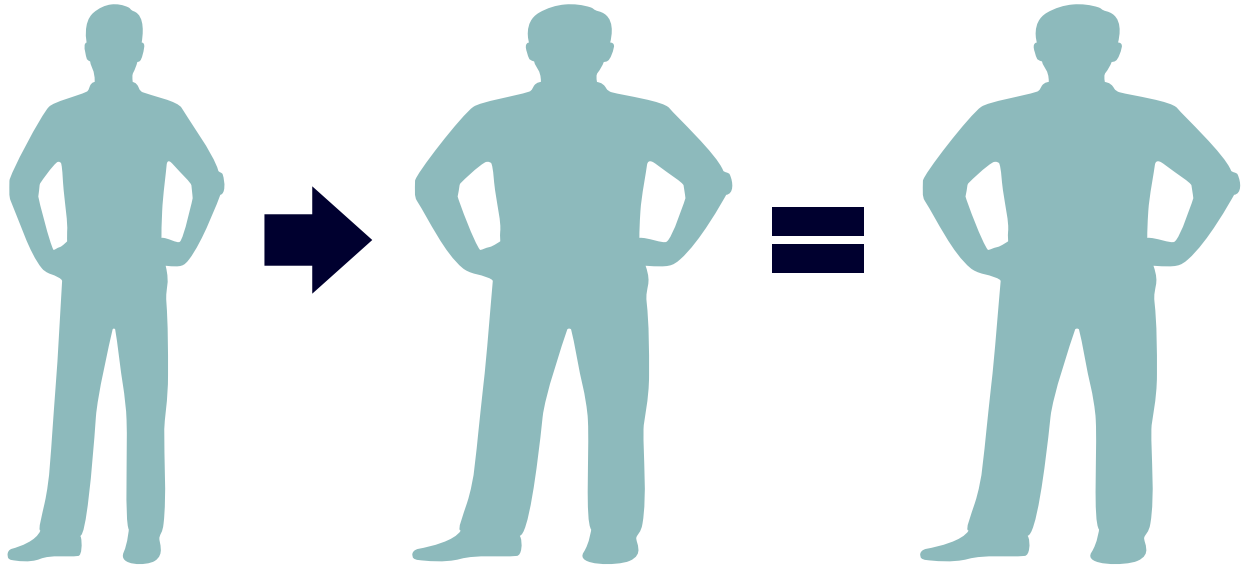


01. Mesh construction & Rigging

7) Boundary matching



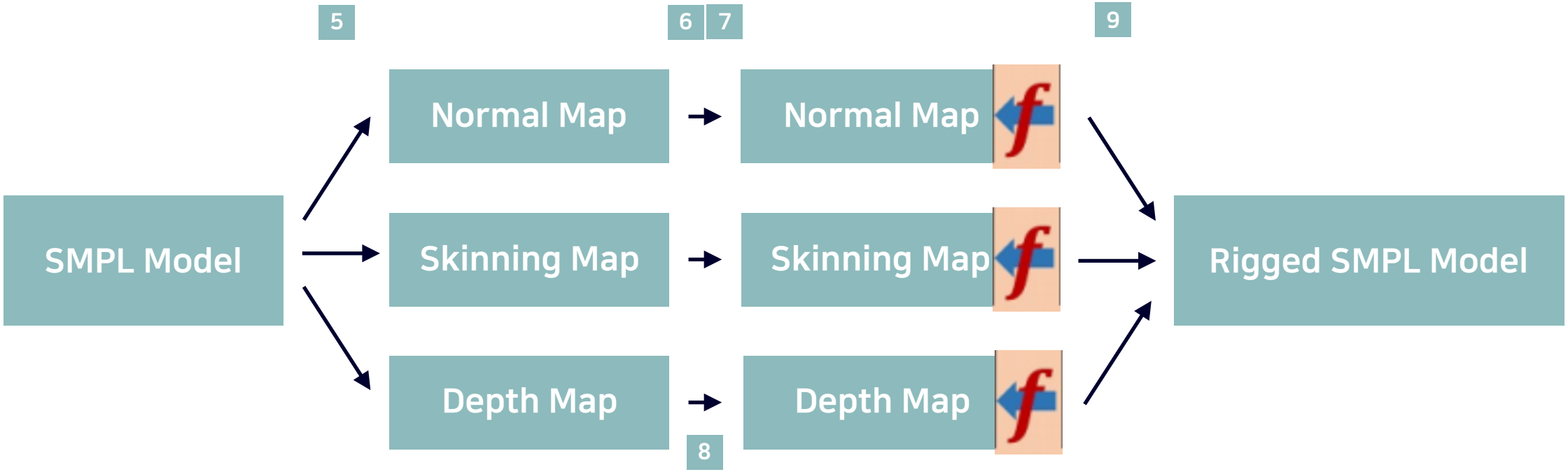
(a) (b)



Hole-filling

01. Mesh construction & Rigging

8) Rebuilding Depth Map



01. Mesh construction & Rigging

8) Rebuilding Depth Map

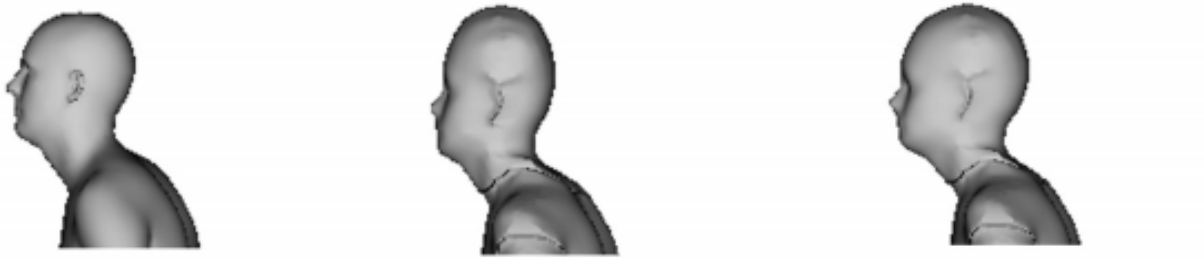
$$S(x) = S_{\text{SMPL}}(f(x)) \tag{1}$$

$$Z_{\partial S}(x \in \partial S) = Z_{\text{SMPL}}(f(x)) \tag{2}$$

$$N(x) = N_{\text{SMPL}}(f(x)) \tag{3}$$

$$Z(x) = \text{Integrate}[N; Z_{\partial S}] \tag{4}$$

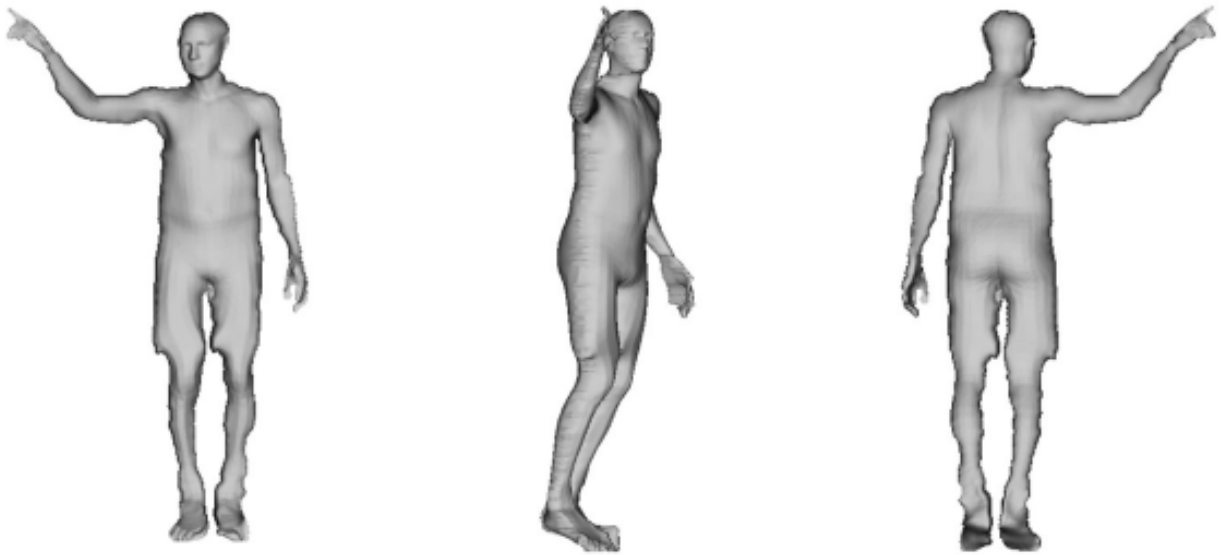
$$W(x) = W_{\text{SMPL}}(f(x)) \tag{5}$$



(a) SMPL (b) Warp depth (c) Warp normal and then integrate

01. Mesh construction & Rigging

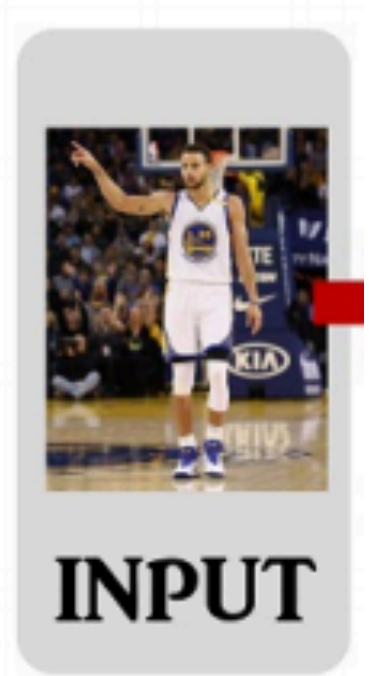
9) Rigged Mesh

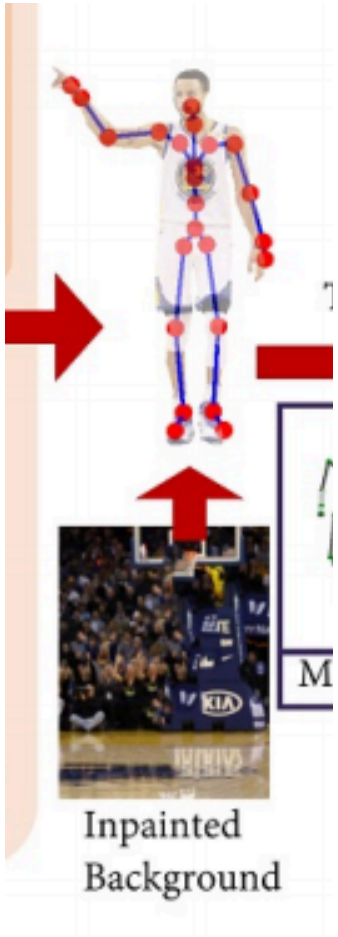


(a) front mesh (b) side view (c) back mesh

01. Mesh construction & Rigging

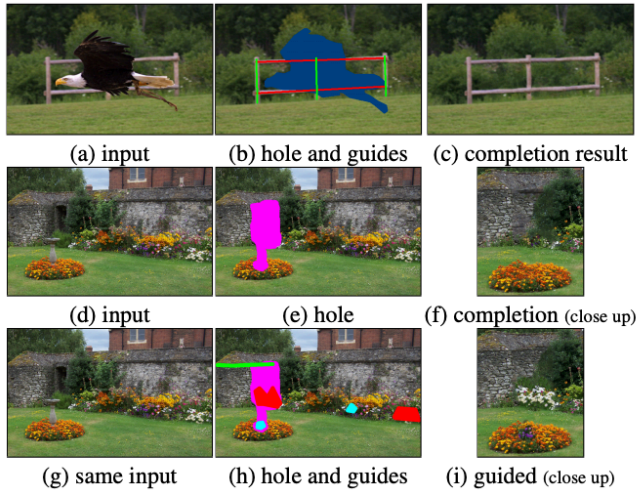
10) Inpainted





(a) original (b) hole+constraints (c) hole filled (d) constraints (e) constrained retarget (f) reshuffle

Figure 1: *Structural image editing. Left to right: (a) the original image; (b) a hole is marked (magenta) and we use line constraints (red/green/blue) to improve the continuity of the roofline; (c) the hole is filled in; (d) user-supplied line constraints for retargeting; (e) retargeting using constraints eliminates two columns automatically; and (f) user translates the roof upward using reshuffling.*



(a) input (b) hole and guides (c) completion result

(d) input (e) hole (f) completion (close up)

(g) same input (h) hole and guides (i) guided (close up)

Figure 4: *Two examples of guided image completion. The bird is removed from input (a). The user marks the completion region and labels constraints on the search in (b), producing the output (c) in a few seconds. The flowers are removed from input (d), with a user-provided mask (e), resulting in output (f). Starting with the same input (g), the user marks constraints on the flowers and roofline (h), producing an output (i) with modified flower color and roofline.*

[14] Barnes, Connelly, et al. "PatchMatch: A randomized correspondence algorithm for structural image editing." ACM Transactions on Graphics (ToG). Vol. 28. No. 3. ACM, 2009.

01. Mesh construction & Rigging

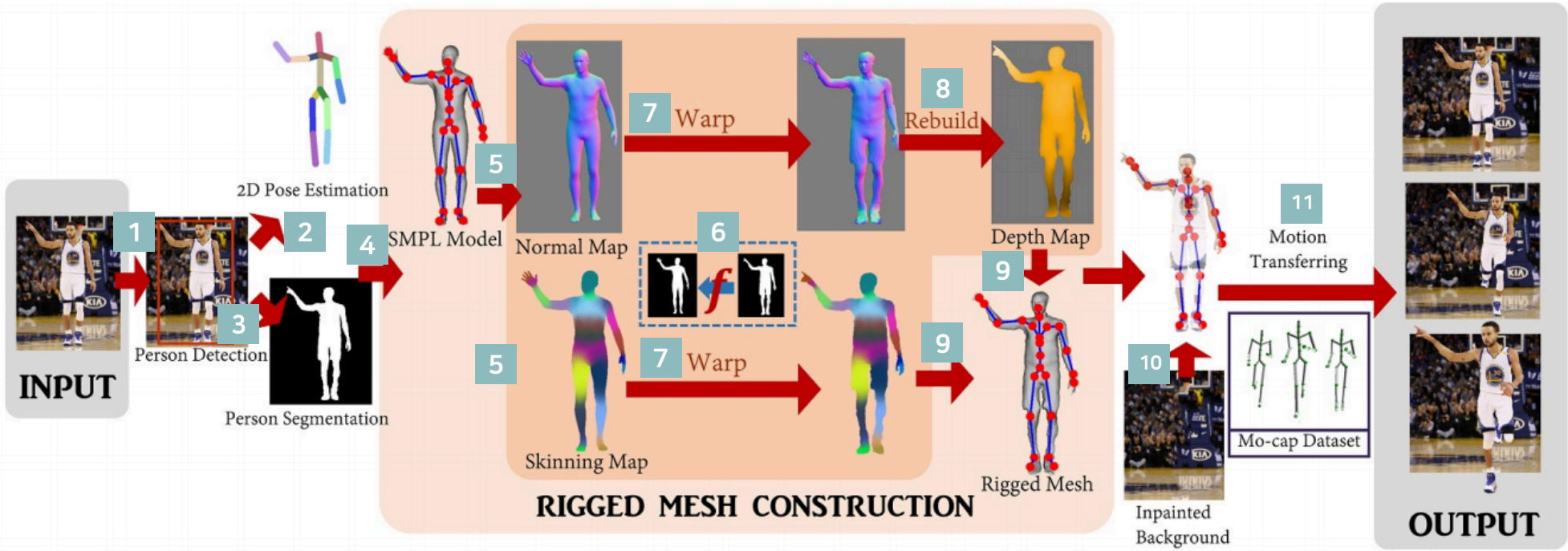
11) Motion transffering

Figure 7: Six animation results. The input is always on left.
Photo credits: [3, 9, 2, 1, 8, 6]

https://youtu.be/G63goXc5MyU

01. Mesh construction & Rigging

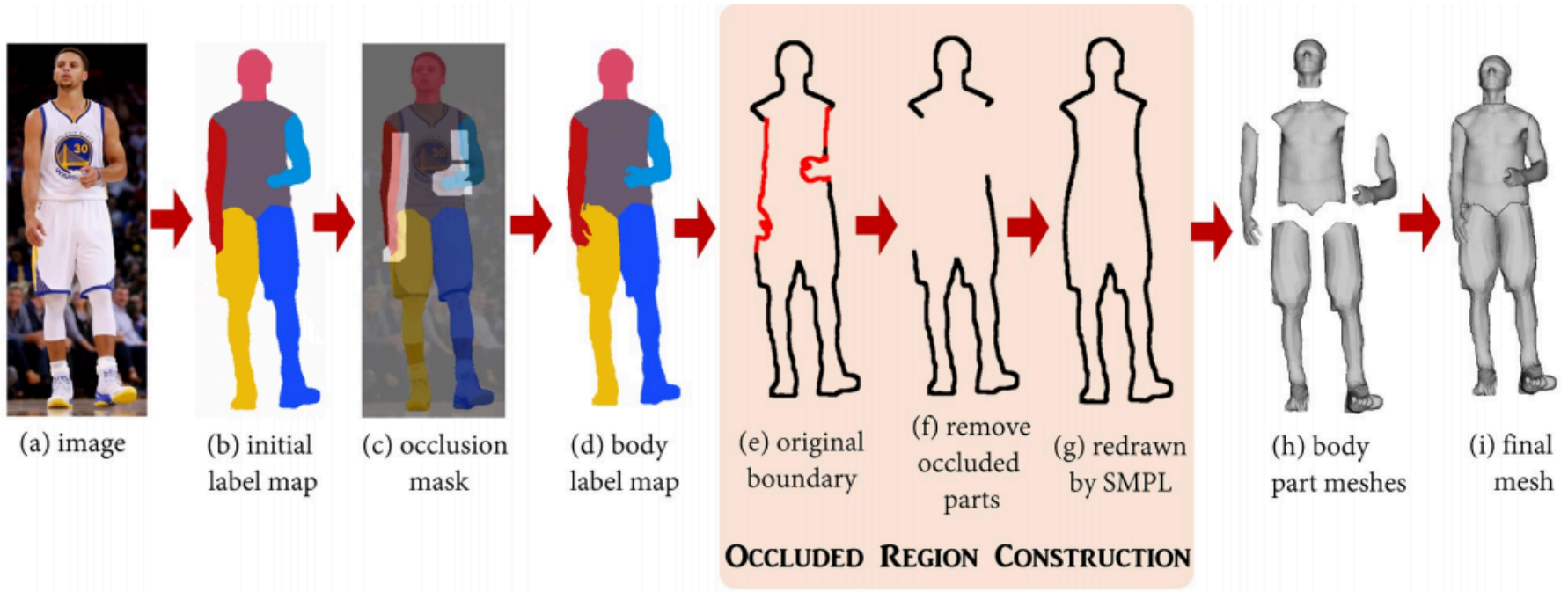
Review



2. Self-occlusion

02. Self-occlusion

self-occlusion?



02. Self-occlusion

(1) Body label map?

(1) estimate an initial label map L(init) for each pixel x to be as similar as possible to L(SMPL)

(2) refine L(init) at occlusion boundaries

02. Self-occlusion

(1) Body label map - initial body labeling

(a) image (b) initial label map

$$\min_{L_{\text{init}}} \sum_{p \in S} U(L_{\text{init}}(p)) + \gamma \sum_{p \in S, q \in \mathcal{N}(p) \cap S} V(L_{\text{init}}(p), L_{\text{init}}(q)) \quad (12)$$

where

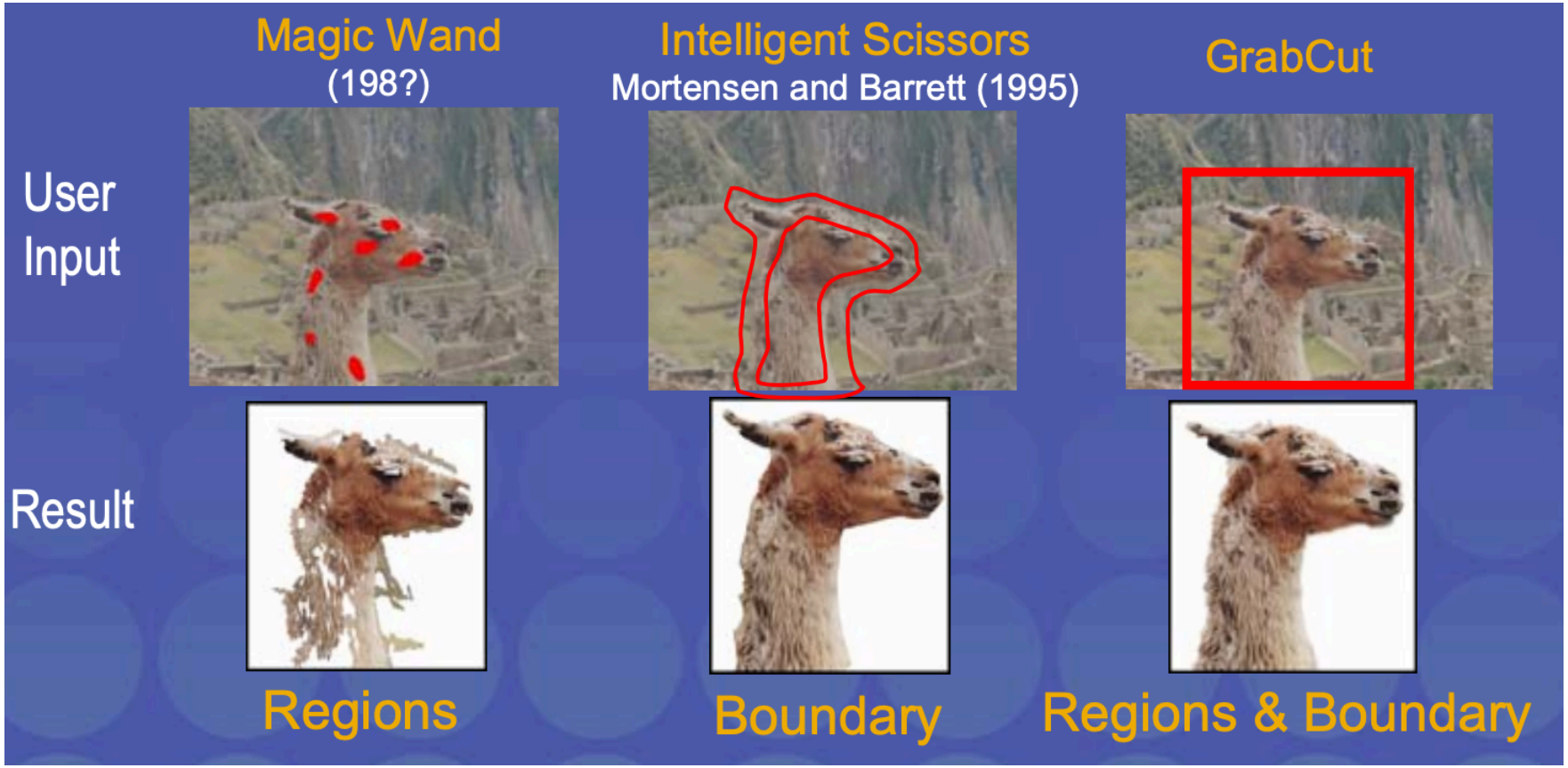
$$U(L_{\text{init}}(p)) = \min_{r | L_{\text{SMPL}}(r) = L(p)} \|p - r\|_2 \quad (13)$$

$$V(L_{\text{init}}(p), L_{\text{init}}(q)) = \begin{cases} 1 & \text{if } L_{\text{init}}(p) \neq L_{\text{init}}(q) \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

[45] Rother, Carsten, Vladimir Kolmogorov, and Andrew Blake. "Interactive foreground extraction using iterated graph cuts." ACM Transactions on Graphics 23 (2004): 3.

02. Self-occlusion

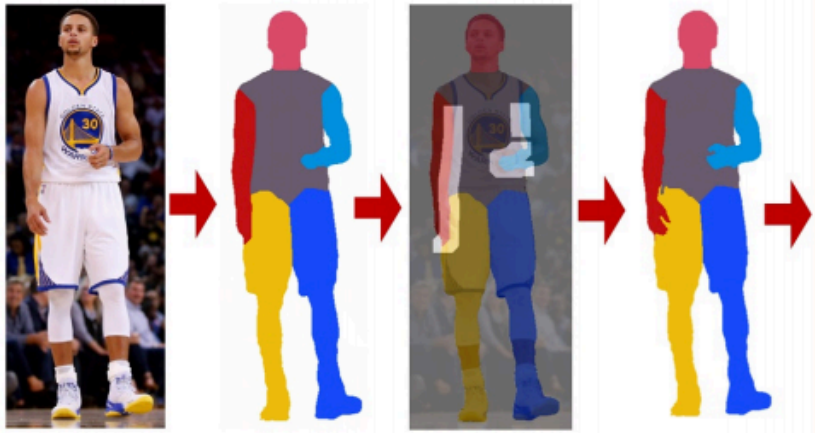
(1) Body label map - refined body labeling



[45] Rother, Carsten, Vladimir Kolmogorov, and Andrew Blake. "Interactive foreground extraction using iterated graph cuts." ACM Transactions on Graphics 23 (2004): 3.

02. Self-occlusion

(1) Body label map - result

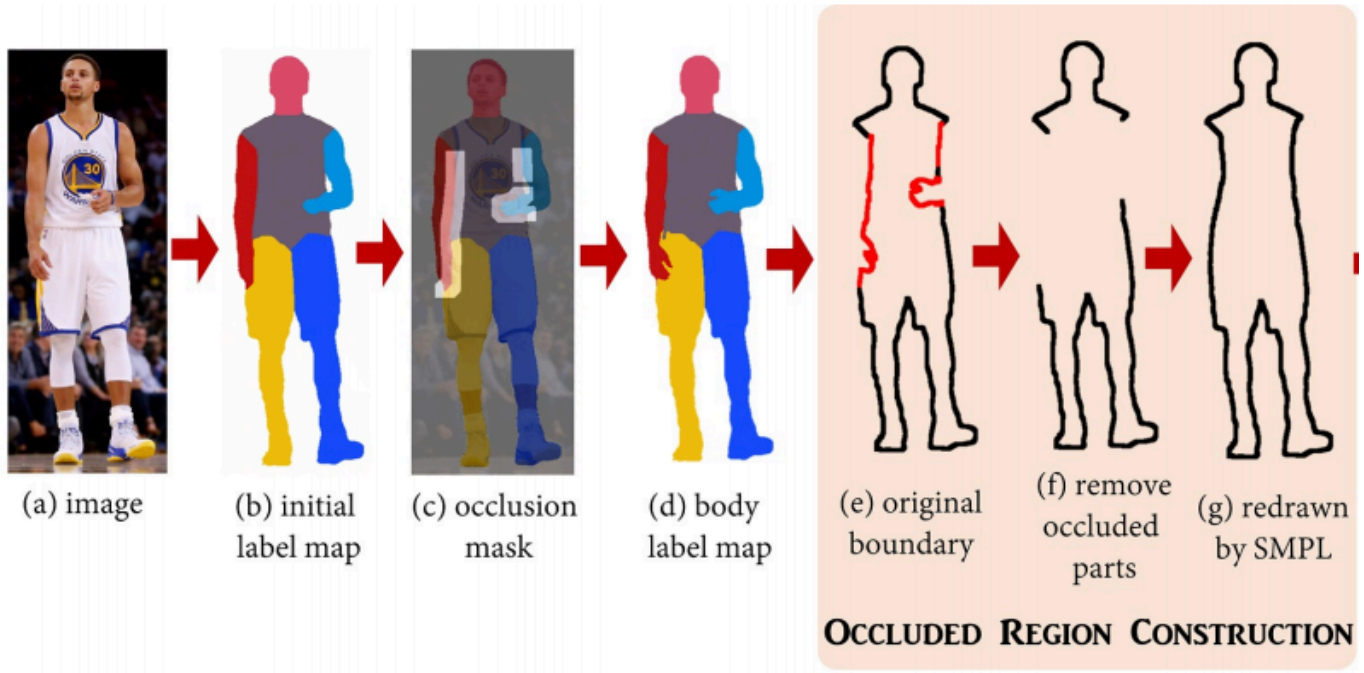


(a) image (b) initial label map (c) occlusion mask (d) body label map

[45] Rother, Carsten, Vladimir Kolmogorov, and Andrew Blake. "Interactive foreground extraction using iterated graph cuts." ACM Transactions on Graphics 23 (2004): 3.

02. Self-occlusion

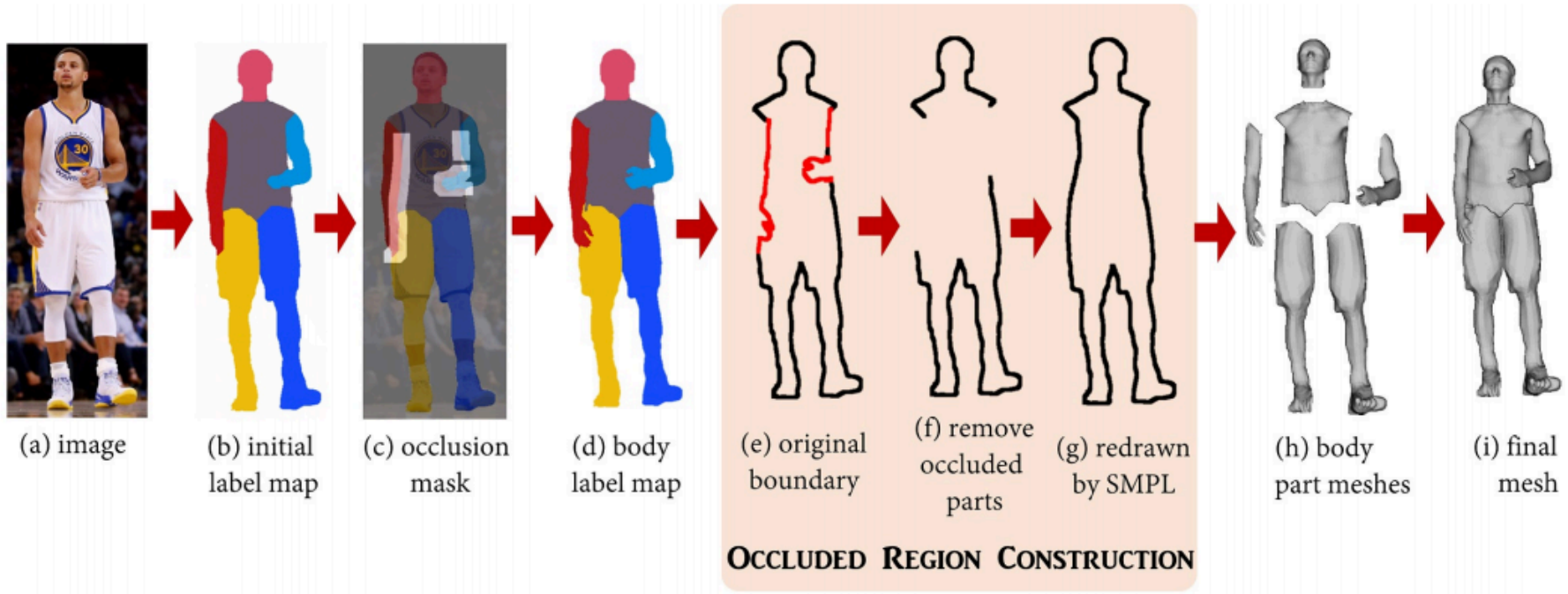
(2) Occluded region construction



[45] Rother, Carsten, Vladimir Kolmogorov, and Andrew Blake. "Interactive foreground extraction using iterated graph cuts." ACM Transactions on Graphics 23 (2004): 3.

02. Self-occlusion

(3) Mesh construction



[45] Rother, Carsten, Vladimir Kolmogorov, and Andrew Blake. "Interactive foreground extraction using iterated graph cuts." ACM Transactions on Graphics 23 (2004): 3.

02. Self-occlusion

review

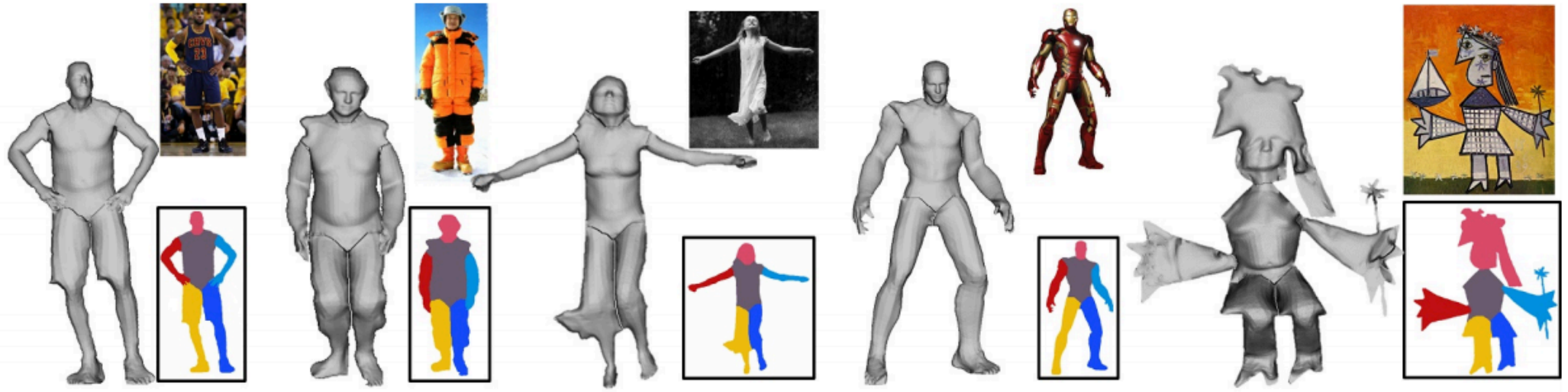


Figure 6: Examples of body label maps and meshes (input photos are put on top right corner). *Photo credits: [3, 5, 4, 6, 8]*

[45] Rother, Carsten, Vladimir Kolmogorov, and Andrew Blake. "Interactive foreground extraction using iterated graph cuts." ACM Transactions on Graphics 23 (2004): 3.

3. Final Steps

03. Final Steps

1) Head pose correction

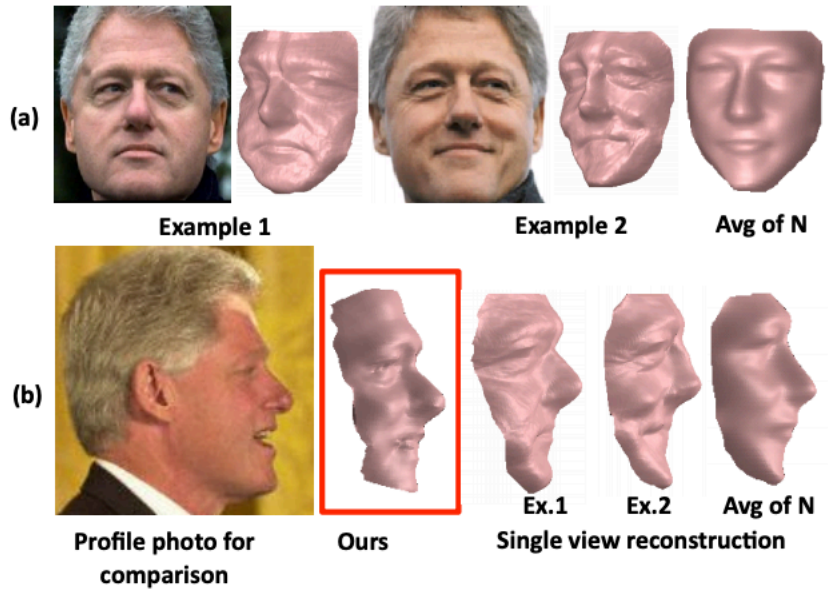


Figure 9. Comparison to single view reconstructions provided by [11]. (a) Two photos of Bill Clinton, the shape reconstruction using the single view method (from each of the images separately), and average of single view reconstructions from all the images in the set. (b) Photo of Clinton's profile, and profile renderings of our reconstruction, two single view reconstructions and average of all the single view reconstructions.

[31] Kemelmacher-Shlizerman, Ira, and Steven M. Seitz. "Face reconstruction in the wild." IEEE, 2011.

03. Final Steps

2) Texturing

a. paste a mirrored copy of the front texture onto the back

b. inpaint with optional user guidance & poison blending

[31] Kemelmacher-Shlizerman, Ira, and Steven M. Seitz. "Face reconstruction in the wild." IEEE, 2011.

03. Final Steps

2) Texturing

a. paste a mirrored copy of the front texture onto the back

[31] Kemelmacher-Shlizerman, Ira, and Steven M. Seitz. "Face reconstruction in the wild." IEEE, 2011.

03. Final Steps

2) Texturing

b. inpaint with optional user guidance & poison blending

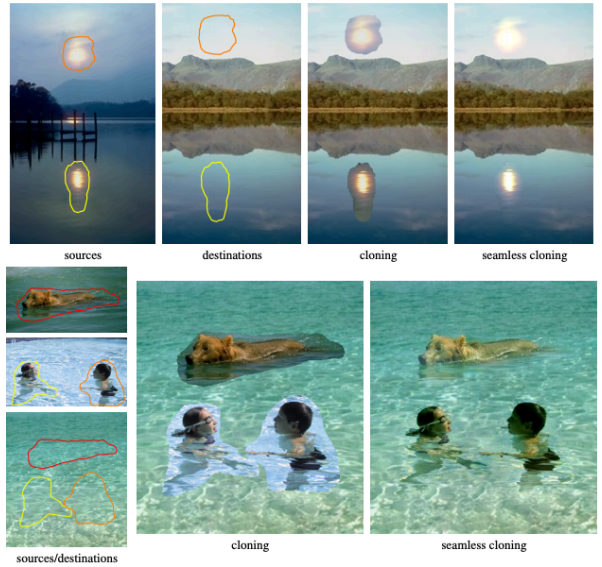


Figure 3: **Insertion**. The power of the method is fully expressed when inserting objects with complex outlines into a new background. Because of the drastic differences between the source and the destination, standard image cloning cannot be used in this case.

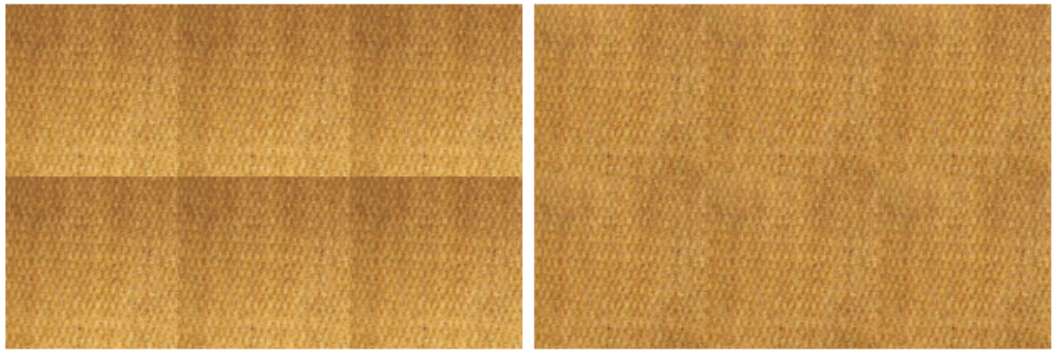


Figure 12: **Seamless tiling**. Setting periodic boundary values on the border of a rectangular region before integrating with the Poisson solver yields a tileable image.

[44] Pérez, Patrick, Michel Gangnet, and Andrew Blake. "Poisson image editing." ACM Transactions on graphics (TOG) 22.3 (2003): 313-318.

04. Results and Discussion

Experience

(a) input photo (b) Hornung et al. (c) ours – 3D demonstration

Figure 10: Comparison result with [26]: (a) input photo; (b) animation method proposed in [26]; (c) 3D demonstration using our method, which is not possible in [26]. *Photo credit: Hornung et al.*

103 응시자
86% 선호

[26] Hornung, Alexander, Ellen Dekkers, and Leif Kobbelt. "Character animation from 2D pictures and 3D motion data." ACM Transactions on Graphics (ToG) 26.1 (2007): 1.

04. Results and Discussion

Limitations

(1) **Shadows and reflections** are currently not modeled by out method and thus won't move with the animation

(2) The shape **may look unrealistic**

(3) Out method accounts for **self-occlusions** when arms **partially occlude** the head, torso, or legs

(4) Person detection and segmentation, pose detection and body labeling **can fail**, requiring manual corrections

(5) We have opted for simple texture inpainting for occluded body parts, with some user interaction if needed

05. Q

and A