

1 **Distortion-Free Wide-Angle Portraits on Camera Phones**

2
3 PENGTAO NI, Carnegie Mellon University, USA

4
5 Wide-angle lenses offer unique perspectives but often introduce undesirable distortions, particularly affecting the portrayal of human
6 subjects. This challenge underscores the importance of our research, which aims to correct these distortions for a more accurate and
7 pleasing representation of individuals in photographs. Building upon Shih et al.'s foundational work in wide-angle portrait correction,
8 this paper presents an innovative approach that extends the scope of correction beyond facial features to include the entire body and
9 torso.

10
11 **ACM Reference Format:**

12 Pengtao Ni. 2023. Distortion-Free Wide-Angle Portraits on Camera Phones. 1, 1 (December 2023), 9 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

36
37 Author's address: Pengtao Ni, pni@andrew.cmu.edu, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, Pennsylvania, USA, 15213.

38
39
40
41
42
43
44 Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not
45 made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components
46 of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to
47 redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

48 © 2023 Association for Computing Machinery.

49 Manuscript submitted to ACM

50
51 Manuscript submitted to ACM

53 **1 INTRODUCTION**

54
 55 The advent of wide-angle lenses in photography has significantly impacted portrait photography, offering creative
 56 perspectives while introducing notable distortions. Shih et al.'s work [2] in wide-angle portrait correction serves as the
 57 foundational basis for our research. Their method focuses on correcting facial distortions caused by these lenses, which
 58 tend to exaggerate features and warp facial proportions, especially when the subject is near the edge of the frame. Shih
 59 et al. developed a content-aware optimization method that identifies facial features in wide-angle portraits and rectifies
 60 the distortions, effectively restoring natural facial proportions. Their method cleverly balances the optimization process
 61 to restore natural facial proportions while maintaining the integrity of the background.

62
 63 Building on their approach to facial correction in wide-angle portraits, our project seeks to extend these principles
 64 to include full-body and torso distortions. By doing so, we aim to address the broader spectrum of distortions that
 65 wide-angle lenses cause, ensuring a more holistic correction for portrait photography.

66 **2 METHOD**

67
 68 Shih et al.'s method begins with detecting face regions in the distorted image and estimating the stereographic projection
 69 from the camera field of view. Next, they optimize a target mesh that performs locally stereographic warp on face
 70 regions while preserving the overall background. Finally, they warp the input image with the optimized mesh to
 71 generate the final result.

72
 73 In this study, we attempt to reimplement Shih et al.'s method as closely as possible. However, due to certain constraints
 74 such as the unavailability of specific methods or resources used in the original study, some adaptations are necessary.
 75 These modifications are made with careful consideration, ensuring that they align with the core principles of the
 76 baseline method while accommodating the limitations encountered. In the following subsections, I will delve into the
 77 details of my implementation, outlining the specific methods and parameters used, and describing how and why they
 78 differ from those in the baseline study.

79 **2.1 Subject Mask Segmentation**

80 To identify face and hair regions, our baseline utilizes a subject segmentation network from the work by Wadhwa et al.
 81 [3]. However, due to access restrictions, we used alternative methods for face detection and subject segmentation.

82
 83 To detect face regions, we utilize MTCNN from the work by Zhang et al. [4], which achieves high accuracy in
 84 detecting facial features across a variety of angles and lighting conditions. Following our baseline's methods, we
 85 empirically expand the bounding-box height and width to include hair and chin.

86 For subject segmentation, we inferece a pretrained Deeplabv3 model with a ResNet-50 backbone [1], which is a
 87 state-of-the-art architecture for subject segmentation tasks.

88
 89 Finally, we take the intersection of the rectangular face bounds and subject mask to obtain the final face mask.

90 **2.2 Stereographic Projection**

91 Stereographic projection is a technique used for mapping a sphere onto a plane. Our baseline leverages stereographic
 92 projection for the facial regions of the wide-angle portraits. By applying this projection, the algorithm adjusts the
 93 distorted facial features to more natural proportions, effectively counteracting the stretching and squishing caused by
 94 the wide-angle lens.

Assuming that the input image is perspectively projected, our baseline computes the stereographic projection using a radial mapping:

$$r_u = r_0 \tan(0.5 \arctan(\frac{r_p}{f})), \quad (1)$$

where f is the camera focal length, r_u and r_p are the radial distances to the optical center under the stereographic and perspective projection, respectively. The scaling factor r_0 is chosen such that $r_u = r_p$ at the image boundary:

$$r_0 = \frac{d}{2 \tan(0.5 \arctan(\frac{d}{2f}))}. \quad (2)$$

Our baseline chooses $d = \min(W, H)$ where W and H are the width and height of the input image, so that $r_u = r_p$ at the narrower image boundaries. However, our experiments indicate that adopting $d = \sqrt{H^2 + W^2}$ yields more effective restoration of distortions, producing results that more closely resembled those of the baseline, as shown below.

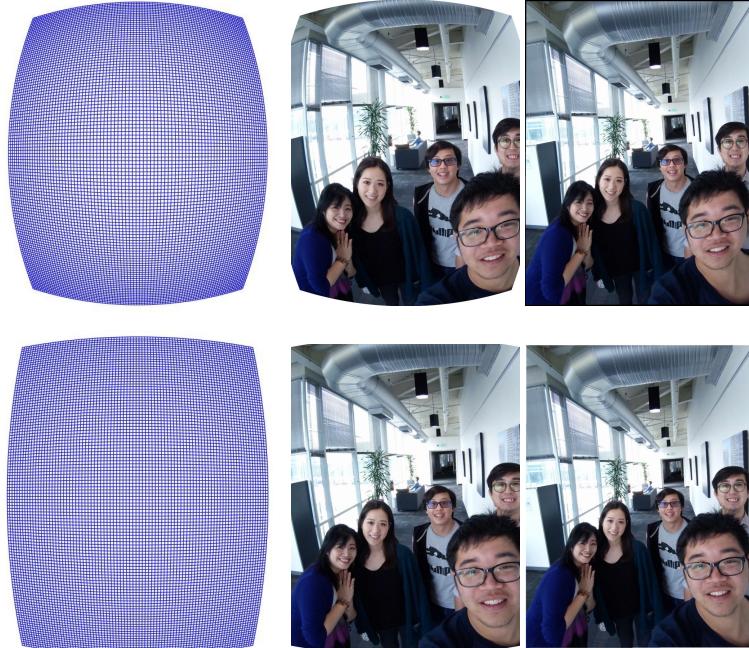


Fig. 1. **Top:** Stereographic mesh, reprojected input, and final output using $d = \min(W, H)$. **Bottom:** Stereographic mesh, reprojected input, and final output using $d = \sqrt{H^2 + W^2}$.

2.3 Mesh Placement

Following our baseline's convention, we denote M for a mesh, which consists of a vertex set $\{\mathbf{v}_i\}$, where i are linearized indices of the 2D coordinates on the grid of M , and \mathbf{v}_i denotes a 2D coordinate.

We assume the input image is projected by perspective projection, and the source mesh is a uniform grid $M_p = \{\mathbf{p}_i\}$. By applying stereographic projection to each \mathbf{p}_i , we create a mesh M_u that represents the input image reprojected by

stereographic projection. We label vertices on faces by sampling the face mask with M_p , and form the face weights $\{w_i\}$:

$$w_i = \begin{cases} 1 & \text{if } p_i \in \text{face mask,} \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

Using an energy minimization approach, we seamlessly transition from M_u at the face regions to M_p at the background.

2.4 Energy Minimization

We minimize the following energy function to determine an optimal mesh $M^* = \{\mathbf{v}_i^*\}$:

$$\{\mathbf{v}_i^*\} = \underset{\mathbf{v}_i}{\operatorname{argmin}} E_t(\{\mathbf{v}_i\}), \quad (4)$$

where E_t is the weighted sum of several energy terms explained below.

Face Objective Term. The face energy E_f encourages each face in the image to follow the stereographic mesh M_u :

$$E_f = \sum_k \sum_{i \in \mathcal{B}_k} w_i m_i \|\mathbf{v}_i - (\mathbf{S}_k \mathbf{u}_i + \mathbf{t}_k)\|_2^2 + \lambda(\mathbf{S}_k), \quad (5)$$

where k indexes the detected faces, \mathcal{B}_k denotes the set of vertices on the k -th face, $\{w_i\}$ are the face weight from Eq. (3), $\{m_i\}$ are the correction strength sampled from a radial sigmoid function to enhance correction for image corners, $\{\mathbf{u}_i\}$ are vertices on the stereographic mesh, \mathbf{S}_k and \mathbf{t}_k represent an affine transform that facilitate the optimizer to find a better solution by slightly translating, rotating and scaling each face individually, and $\lambda(\mathbf{S}_k)$ is a regularization term to preserve the face scale.

Line-Bending Term. The line-bending term E_b preserves straight lines in the background by encouraging the output mesh to scale rather than twist:

$$E_b = \sum_i \sum_{j \in N(i)} \|(\mathbf{v}_i - \mathbf{v}_j) \times \mathbf{e}_{ij}\|_2^2, \quad (6)$$

where $N(i)$ is the set of neighboring indices of i , \mathbf{e}_{ij} is the unit vector along the direction $\mathbf{p}_i - \mathbf{p}_j$, and \times denotes the cross product.

Regularization Term. The regularization term E_r encourages smoothness between 4-way adjacent vertices:

$$E_r = \sum_i \sum_{j \in N(i)} \|\mathbf{v}_i - \mathbf{v}_j\|_2^2. \quad (7)$$

Asymmetric Cost Term. The asymmetric cost term E_a reduces undefined regions after mesh warping by encouraging the vertices on the mesh boundaries to expand instead of shrinking:

$$E_a = E_l + E_r + E_t + E_b, \quad (8)$$

$$E_l = \mathbb{I}(v_{i,x} > 0) \cdot \|v_{i,x}\|_2^2, \forall i \in \delta_{\text{left}},$$

$$E_r = \mathbb{I}(v_{i,x} < W) \cdot \|v_{i,x} - W\|_2^2, \forall i \in \delta_{\text{right}},$$

$$E_t = \mathbb{I}(v_{i,y} > 0) \cdot \|v_{i,y}\|_2^2, \forall i \in \delta_{\text{top}},$$

$$E_b = \mathbb{I}(v_{i,y} < H) \cdot \|v_{i,y} - H\|_2^2, \forall i \in \delta_{\text{bottom}}, \quad (9)$$

where $\mathbb{I}(\cdot)$ is the indicator function that returns 1 for true condition and 0 otherwise, and δ_* are the mesh boundaries.

Boundary Condition. Our baseline first expands the mesh domain by padding q vertices on each side, and then enforce the following boundary conditions on the extended mesh:

$$\begin{cases} v_{i,x} = p_{i,x}, \text{ if } i \in \text{left or right boundary,} \\ v_{i,y} = p_{i,y}, \text{ if } i \in \text{top or bottom boundary.} \end{cases} \quad (10)$$

That is, the boundary vertices can only move along borders.

2.5 Optimization

The total energy function in the baseline is the weighted combination of the energy terms from Eqs. (5), (6), (7), and (8):

$$E_t = \lambda_f E_f + \lambda_b E_b + \lambda_r E_r + \lambda_a E_a, \text{ subject to Eq. (10),} \quad (11)$$

Alternatively, we introduce a energy term for the boundary condition:

$$E_{bc} = E_x + E_y, \quad (12)$$

$$E_x = \|v_{i,x} - p_{i,x}\|_2^2, \forall i \in \delta_{\text{left}} \text{ or } i \in \delta_{\text{right}},$$

$$E_y = \|v_{i,y} - p_{i,y}\|_2^2, \forall i \in \delta_{\text{top}} \text{ or } i \in \delta_{\text{bottom}}, \quad (13)$$

and the total energy function becomes an unconstrained one:

$$E_t = \lambda_f E_f + \lambda_b E_b + \lambda_r E_r + \lambda_a E_a + \lambda_{bc} E_{bc}, \quad (14)$$

where $\lambda_f, \lambda_b, \lambda_r, \lambda_a$, and λ_{bc} are the weights for the corresponding energy terms.

2.6 Torso Correction

Building upon our baseline, we explore the correction of entire bodies in addition to just faces. This extension involves substituting face weights with body weights, which were derived from sampling the subject mask generated through subject segmentation:

$$w_i = \begin{cases} 1 & \text{if } p_i \in \text{subject mask,} \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

Given that body correction impacts considerably larger areas than face correction alone, it becomes necessary to reduce the energy weight assigned to the boundary condition. This reduction allows for greater deformation at the image boundaries, accommodating the more extensive correction needs of full body adjustment. Detailed discussions of the specific parameter adjustments and the corresponding results are elaborated in Section 3.2.

2.7 Implementation Details

We set the mesh dimension before padding to 103×78 for all of our experiments. We implement and test our solution on desktop with an AMD Ryzen 5 3400G CPU. For a 2K image, the MTCNN face detector runs for 5 seconds on average, and inferencing the pretrained Deeplabv3 model takes 2 minutes on average.

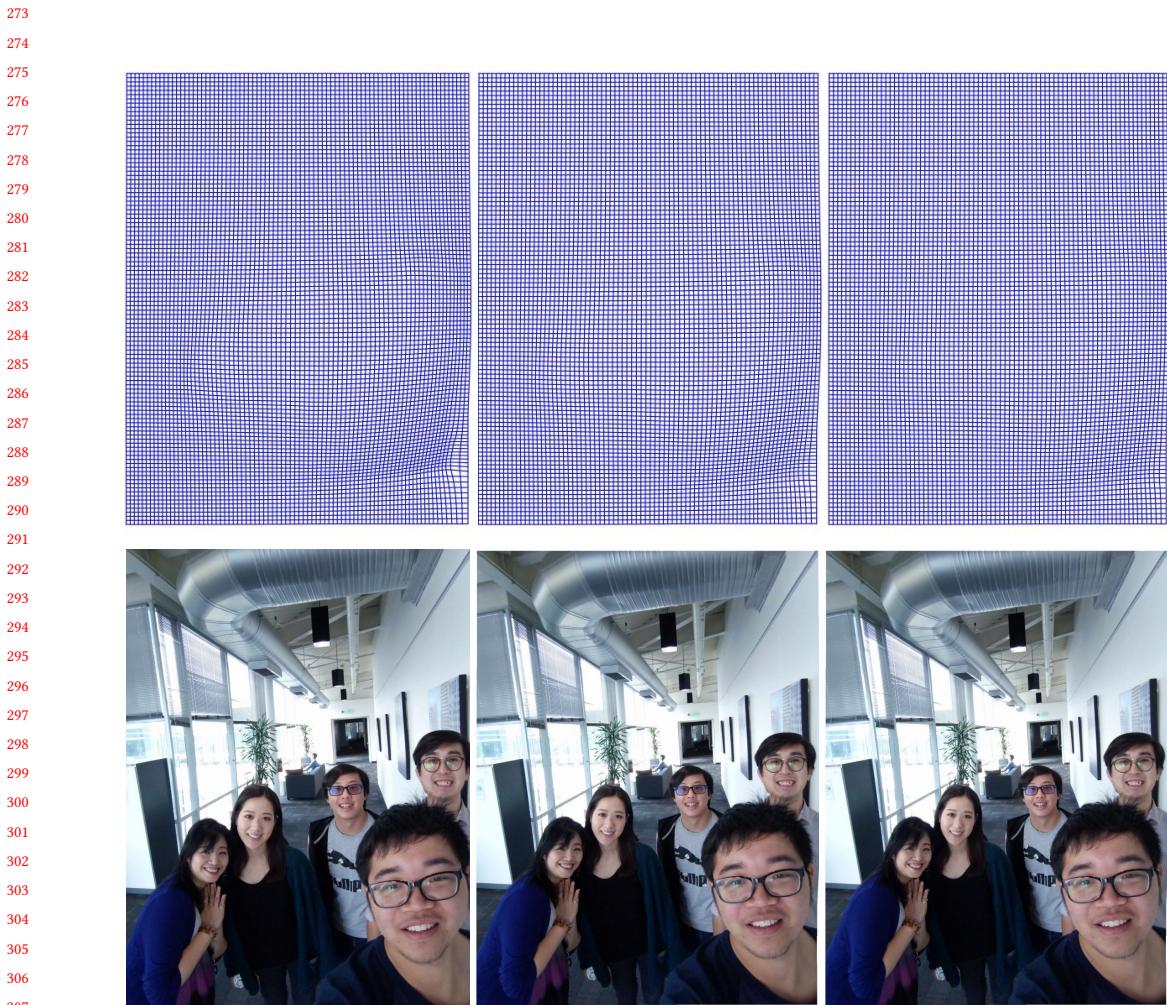
We initialize the mesh as the uniform source mesh and solve the least-squared problem in Eq. (14) with PyTorch. We use the Adam optimizer with 1.0 learning rate. We run the optimization for 200 epochs, which takes under 5 seconds on average.

261 **3 RESULTS**

262 **3.1 Ablations**

264 Due to variations in implementation specifics, particularly our selection of d for the stereographic projection, we found
 265 it necessary to adjust several parameters, including padding size and energy weights, to achieve results comparable to
 266 the baseline study. The outcomes of these adjustments are demonstrated in the figure below.

267 Initially, we notice unnatural distortions in the face near the bottom of the right boundary, prompting us to increase
 268 the padding to relax these boundary conditions. Following this, we enhance the weights for both the line-bending and
 269 regularization terms. The adjustment in the line-bending term was aimed at preserving the straightness of lines, while
 270 the increase in the regularization term's weight was focused on maintaining the smoothness of the background.
 271



308 Fig. 2. **Left:** Using the baseline's parameters. **Middle:** Increasing padding from 4 to 6. **Right:** Increasing λ_b and λ_r to 16.0 and 4.0.

313 3.2 Torso Correction

314
 315 We explore the correction of entire bodies instead of just faces by substituting face weights with body weights. Due
 316 to the significantly larger areas impacted by body correction, we find it necessary to lower the energy weight for the
 317 boundary condition, reducing it from $1e8$ to 1. Additionally, to mitigate the issue of undefined regions commonly
 318 associated with larger correction areas, we increase the energy weight for the asymmetric cost term to favor expansion
 319 over shrinking during the correction process. The results are illustrated below.
 320



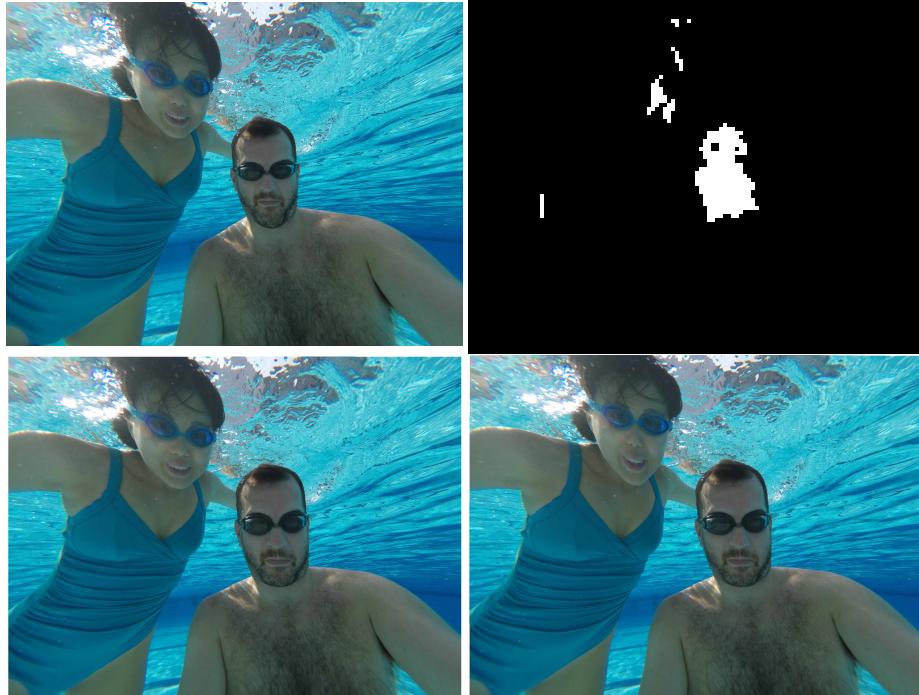
358 Fig. 3. **Left:** Input image. **Middle:** Face correction result. **Right:** Face and torso correction result (cropped).
 359

360 The results clearly demonstrate effective correction of human torsos to a more natural appearance. However, it's
 361 noteworthy that the warping occurring at the boundaries necessitates cropping a slightly larger area of the image to
 362 maintain the desired visual integrity post-correction.
 363

365 3.3 Challenging Cases

366 Erroneous subject segmentation fails in correction as illustrated by the example below.

**367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394**



**395 Fig. 4. Top-left: Input image. Top-right: Subject mask. Bottom-left: Face correction result. Bottom-right: Face and torso correction
396 results.**

401 4 CONCLUSION

In this project, we reimplement Shih et al.'s work with PyTorch and achieved comparable results. We also present a novel approach to extending the wide-angle portrait correction method, focusing on the comprehensive correction of entire bodies in addition to faces.

Our results indicate a significant improvement in the naturalness of body proportions in wide-angle photographs. However, a notable trade-off in our approach is the increased need for cropping due to warping at the image boundaries. This aspect highlights the complex balance between distortion correction and image composition, especially in the context of wide-angle photography.

Future work could focus on refining the balance between correction effectiveness and the preservation of the original image's composition. Additionally, further enhancing the subject segmentation techniques could further improve the usability and adaptability of our method for a broader range of wide-angle images.

417 REFERENCES

- 418 [1] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. 2017. Rethinking atrous convolution for semantic image segmentation.
419 *arXiv preprint arXiv:1706.05587* (2017).
- 420 [2] YiChang Shih, Wei-Sheng Lai, and Chia-Kai Liang. 2019. Distortion-free wide-angle portraits on camera phones. *ACM Transactions on Graphics
(TOG)* 38, 4 (2019), 1–12.
- 422 [3] Neal Wadhwa, Rahul Garg, David E Jacobs, Bryan E Feldman, Nori Kanazawa, Robert Carroll, Yair Movshovitz-Attias, Jonathan T Barron, Yael Pritch,
423 and Marc Levoy. 2018. Synthetic depth-of-field with a single-camera mobile phone. *ACM Transactions on Graphics (ToG)* 37, 4 (2018), 1–13.
- 424 [4] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. 2016. Joint face detection and alignment using multitask cascaded convolutional networks.
425 *IEEE signal processing letters* 23, 10 (2016), 1499–1503.

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

460

461

462

463

464

465

466

467

468