Distortion-Free Wide-Angle Portraits on Camera Phones

YICHANG SHIH, WEI-SHENG LAI, and CHIA-KAI LIANG, Google



(a) A wide-angle photo with distortions on subjects' faces.

(b) Distortion-free photo by our method.

Fig. 1. (a) A group selfie taken by a wide-angle 97° field-of-view phone camera. The perspective projection renders unnatural look to faces on the periphery: they are stretched, twisted, and squished. (b) Our algorithm restores all the distorted face shapes and keeps the background unaffected.

Photographers take wide-angle shots to enjoy expanding views, group portraits that never miss anyone, or composite subjects with spectacular scenery background. In spite of the rapid proliferation of wide-angle cameras on mobile phones, a wider field-of-view (FOV) introduces a stronger perspective distortion. Most notably, faces are stretched, squished, and skewed, to look vastly different from real-life. Correcting such distortions requires professional editing skills, as trivial manipulations can introduce other kinds of distortions. This paper introduces a new algorithm to undistort faces without affecting other parts of the photo. Given a portrait as an input, we formulate an optimization problem to create a content-aware warping mesh which *locally* adapts to the stereographic projection on facial regions, and seamlessly evolves to the perspective projection over the background. Our new energy function performs effectively and reliably for a large group of subjects in the photo. The proposed algorithm is fully automatic and operates at an interactive rate on the mobile platform. We demonstrate promising results on a wide range of FOVs from 70° to 120°.

CCS Concepts: • **Computing methodologies** → **Computational photography**; **Image processing**.

Additional Key Words and Phrases: Content-Aware Warping, Computational Photography, Perspective Correction.

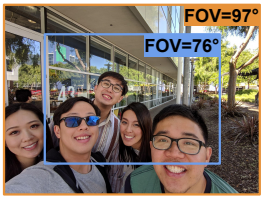
Authors' address: YiChang Shih, yichang@google.com; Wei-Sheng Lai, wlai24@ucmerced.edu; Chia-Kai Liang, ckliang@google.com, 1600 Amphitheatre Parkway, Mountain View, CA, Google.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
© 2019 Copyright held by the owner/author(s).
0730-0301/2019/7-ART61
https://doi.org/10.1145/3306346.3322948

ACM Reference Format:

YiChang Shih, Wei-Sheng Lai, and Chia-Kai Liang. 2019. Distortion-Free Wide-Angle Portraits on Camera Phones. *ACM Trans. Graph.* 38, 4, Article 61 (July 2019), 12 pages. https://doi.org/10.1145/3306346.3322948

1 INTRODUCTION



Empowered with extra peripheral vision to "see", a wide-angle lens is incredibly capable of capturing commemorative moments filled with more people and landscapes, and has been widely used for wedding, sport, landscape, and street photography. For mobile phones, as illustrated in the thumbnail at right, zooming out from normal to wide FOV enables group selfies with friends and families using a handheld camera. Adopting wide-angle cameras has been a recent trend among premium phones. For example, LG G6 has a 100° FOV front camera and a 125° FOV rear camera.

Unfortunately, a wide-angle lens distorts faces when projecting the surrounding world onto a flat image. It leads to unnatural, wider, asymmetric, and unpleasant faces as shown in the inset above and Fig. 1a, and gives misleading impressions of the subjects. In the modern mobile era, people take and share portrait shots, selfies, and group selfies by phones all the time [Izadinia et al. 2015]. However, in our study, all wide-angle camera phones suffer from either perspective distortion, or fish-eye like artifacts that bend straight edges on buildings, facades, interiors, and window frames.

We present an automatic algorithm to reverse perspective distortion on portraitures, so that everyone in the photo looks natural



(a) Input (perspective projection) (b) Stereographic projection (c) Mercator projection (d) Our method

Fig. 2. Applying global projections to the input (a) creates fish-eye artifacts that bend the straight lines in the scene (b)(c). Our method recovers the face shape from distortion *and* preserves the background without bending straight edges (d).

and real. Given an input image, we compute the subject mask to assign per-vertex weights on a coarse mesh over the input image. Then, the core of our approach formulates energy terms that encourage facial vertices to locally emulate the stereographic projection, a conformal mapping between a sphere and a plane, for distortion restoration. Our output combines both the stereographic and perspective projections on a single image. The energy function encourages smooth transitions between the two conflicting projections at face boundary. Different from existing works in generic perspective correction [Carroll et al. 2009; Kopf et al. 2009; Lee et al. 2012], our method is designed specifically to address human faces for its unique challenges and importance for mobile and consumer cameras.

We validate our algorithm by demonstrating image quality improvement on more than four thousand portraits, with 1 to 10+ subjects, taken by cameras ranging from 70° to 120° FOV, which include normal to ultra-wide-angle lenses equipped with major mobile phones. It is crucial for mobile users to immediately see and share captured photos. Delivering a practical solution under these constraints requires novel systematic and algorithmic designs. Our method is highly efficient in memory, CPU, and GPU usage, and processes a 12-megapixel image in 920ms on a modern phone. The process is often finished before users seeing the photos.

We make the following contributions in this work:

▷ We propose an automatic and efficient algorithm to correct perspective distortion on facial regions for wide-angle portrait photography.

▷ We introduce a novel face objective term to incorporate the stereographic and perspective projections on faces and background, respectively.

2 RELATED WORK

We review the methods in camera model, 2D image warping, and face enhancement. As these topics are extensively studied, we discuss the most relevant ones to our work.

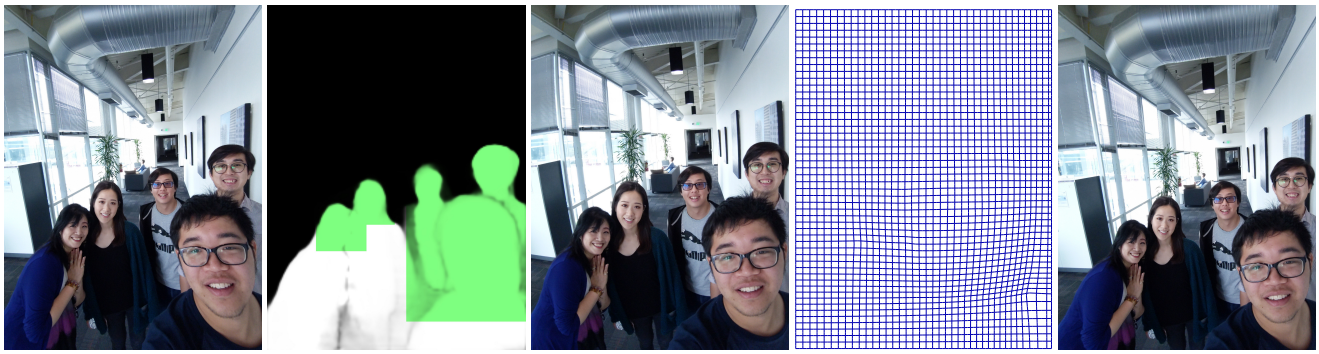
3D Projection. 3D projection inevitably introduces distortions to the rendered 2D image [Zorin and Barr 1995]. The perspective projection distorts the perceived geometry for objects far away from the camera center, especially when the camera FOV is wider than the comfort zone of our foveal vision system [Vishwanath et al. 2005]. Artists dated to early Renaissance already discover the perspective distortion and refer to it as "anamorphosis" [Veltman 1986]. They cautiously handle wider paintings such as architectural scene when the FOV exceeds 60° [Sharpless et al. 2010].

Taking a group photo with a handheld camera often requires wider FOV than 60°. For wide-angle imageries *e.g.*, GoPro camera [GoPro 2017] and panoramas, the stereographic, Mercator, and Pannini projections are widely used to alleviate perspective distortion [Sharpless et al. 2010; Snyder 1987]. These global projections have side effects that bend long and salient straight edges common in human-made structures, resulting the loss of photorealism (Fig. 2). Our method uses a locally adaptive mesh to avoid the line-bending artifacts.

Lens Distortion. The lens distortion is another factor that causes the artifacts in the rendered images. Lens distortion is attributed to the lens design process and difficult to avoid for wide-angle lens. It usually distorts the straight lines at image corners [Brown 1966]. The distortion profile can be calibrated [Zhang 2000] and numerous methods can address this issue [Abràmoff et al. 2004].

However, lens distortion corrected images still exhibit perspective distortion. Portrait photographers have to use telephoto lens, or carefully guide the subjects to the camera center. Our method corrects faces anywhere in the camera view, and sets portrait photographers free from these composition restrictions.

Perspective Distortion Manipulation. Since projection center is free from distortion, we can reduce the distortion in post-processing by carefully selecting a new imaging plane with a new virtual camera view or the planar projection geometry [Pavić et al. 2006; Tehrani et al. 2016]. Existing methods apply a global homography warping determined either manually, such as Perspective Warp feature in Photoshop [Intwala and Agarwala 2015], or automatically [Lee et al. 2012]. Instead, our work performs local corrections to preserve the viewing angle and FOV of the original shot.



(a) Input (97° FOV) (b) Subject mask of (a) (c) Stereographic projection (d) Optimized mesh (e) Distortion-free output

Fig. 3. Overview of our approach: given a wide-angle input rendered with the perspective projection (a), we first retrieve the facial regions using a person segmentation (b), and compute the stereographic projection (Sec. 3). In (b), we mark the region intersecting with the extended face boxes with green color. In (c), we visualize the stereographic projection by warping (a). Then we create a distortion correction mesh (d) using an energy optimization for face unwarping (Sec. 4). Finally, we warp (a) using (d) to create the final output (e). In (e), the face at bottom-right restores the distortion from (a) and looks natural, and the background avoids the fish-eye distortion from (c).

Several mesh-based methods require users to provide scene constraints such as straight lines or vanishing points, and create results by minimizing distortion metrics like conformality cost [Carroll et al. 2010, 2009; Kanamori et al. 2011]. In contrast, our algorithm is fully automatic and executes at interactive rate on the phone for instant sharing.

Content-Aware Warping. Our method belongs to content-aware warping, which has been applied in various image manipulations: panorama stitching and reshaping [Agarwala et al. 2006; Chang et al. 2014; Chen and Chuang 2016; He et al. 2013; Kopf et al. 2009; Li et al. 2015; Zelnik-Manor et al. 2005], correcting wide-angle images [Carroll et al. 2009; Chang and Chuang 2012] and videos [Wei et al. 2012], image and video retargeting [Chang et al. 2011; Wang et al. 2008; Wolf et al. 2007], texture deformation [Gal et al. 2006], stereoscopic editing [Chang et al. 2011; Du et al. 2013], and video stabilization [Guo et al. 2016; Liu et al. 2009, 2011; Zhang et al. 2016]. Different from existing methods that restore geometries of generic objects, our work specifically addresses portrait photos. Our face-specific problem is uniquely challenging when multiple faces cluster together, since human eyes are highly sensitive to artifacts on faces.

Face Undistortion. Using 3D face models, Fried et al. manipulate the camera focal length to correct foreshortening distortion in portraits [2016]. Our work enhances portraits taken in the wild, such as group photos with arbitrary number of subjects, face poses, and occlusions. In these cases, acquiring accurate and robust face models is difficult. Foreshortening happens when faces are very close to the camera, and is a different face distortion problem from our work.

Different from face beautification that relies on the statistical consensus about beauty [Leyvand et al. 2008; Zhang et al. 2011], our goal is to make faces at sides of a photo look like as if they were taken from the camera center with minimal distortion. To achieve this, we use a combination of two different projective geometries on a single image without relying on the aesthetic preferences.

3 PRELIMINARIES AND OVERVIEW

Fig. 3 shows an overview of the proposed algorithm. We assume that the input to our method (Fig. 3a) is perspectively projected. We first identify face regions by computing a subject segmentation mask (Fig. 3b), and estimate the stereographic projection from the camera focal length (Fig. 3c). Then, we compute a target mesh that performs locally stereographic warp on face regions (Fig. 3d) using an energy minimization (Sec. 4). This step outputs an optimized mesh that corrects faces without alternating the overall look of the background. To generate the final result (Fig. 3e), we warp the input with the optimized mesh to undo the distortion on faces.

Subject Mask Segmentation. Our goal is to correct face and hair regions for all the subjects. To identify these areas, we compute the mask of faces and hair using an off-the-shelf subject segmentation network from the work by Wadhwa et al. [2018], as an example illustrated in Fig. 3b. To concentrate the correction on faces and hair, we generate a face mask by intersecting the subject mask with rectangular face bounds returned from face detectors. To ensure we cover the hair and chin, we extend the face bounds by retrieving 128 facial landmarks over the facial contour, including chin, and expand the original face bounding-box to cover all the landmarks. To include hair, we empirically extend the box height by 2 times along the top direction, and the width by half on both sides. We show the trimmed face mask with a green-color-coded map in Fig. 3b. In our experiments, the face mask is robust to face pose, occlusion, large group, and low-light conditions.

Stereographic Projection. Stereographic projection can picture the 3D world onto a 2D plane with minimal conformal distortion, at the expense of changing the curvature of long lines [Carroll et al. 2009; Zorin and Barr 1995]. Fig. 2b and Fig. 3c show that the stereographic projection corrects local objects such as faces.

We emulate the stereographic projection locally to correct perspective distortion for faces. Given the camera focal length f, we compute the stereographic projection from the input using a radial mapping [Zorin and Barr 1995]:

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

where 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:

$$r_0 = \frac{d}{2\tan(0.5 \arctan(\frac{d}{2f}))}, \tag{2}$$

where $d = \min(W, H)$ is the width and height of the input image.

Mesh Placement. We use the bold font for column vectors: $\mathbf{c} = [c_x, c_y, \ldots]^\top$, and denote M for a mesh, *i.e.* a connected graph of grid topology, as an example illustrated in Fig. 3d. A mesh M 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. Given a mesh M defined on the input domain, our warp is defined as a dense 2D vector-to-vector mapping by interpolating $\{\mathbf{v}_i\}$ in M.

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 Eq. (1) to each $\mathbf{p}_i$, we create a mesh M_u that represents the input image reprojected by stereographic projection. Fig. 3c shows the image by warping M_p to M_u. We can see the the face distortion is reduced but the background edges are curved.

A trivial solution is interpolation between M_p and M_u adaptively. We label vertices on faces by sampling the face mask with M_p, and form the face weights denoted as $\{w_i\}$:

$$w_i = \begin{cases} 0 & \text{if } \mathbf{p}_i \notin \text{face mask}, \\ 1 & \text{if } \mathbf{p}_i \in \text{face mask}. \end{cases} \tag{3}$$

As demonstrated in Fig. 4, this solution can show severe artifacts due to large displacements between these two meshes. Simple mask manipulation would not improve the quality, as propagating M_u outside the faces would introduce background distortion. Therefore, we propose an energy minimization approach to reconcile the conflicts between these two projections at face boundaries in the next section.

4 LOCAL FACE UNDISTORTION

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

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

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



(a) Input (97° FOV) (b) Inset of (a) (c) Naive blending (d) Our method

Fig. 4. A wide-angle image (a) distorts the face (b) at the corner. (c) Naively replicating the stereographic mesh M_u from face regions onto the uniform grid M_p creates disturbing artifacts around the face boundary. (d) Our optimization in Sec. 4 resolves the artifacts between the face boundaries.



(a) Input (b) w/o $\mathbf{S}_k, \mathbf{t}_k$ in Eq. (6) (c) With $\mathbf{S}_k, \mathbf{t}_k$ in Eq. (6)

Fig. 5. Given a group portrait with multiple faces close to each other (a), the result without hidden variables $\mathbf{S}_k$ and $\mathbf{t}_k$ in Eq. (6) creates conflicting artifacts visible on faces (b). Our method uses hidden variables to output natural looks for all the subjects (c). (a) is cropped from the top-left corner of a 103° FOV photo.

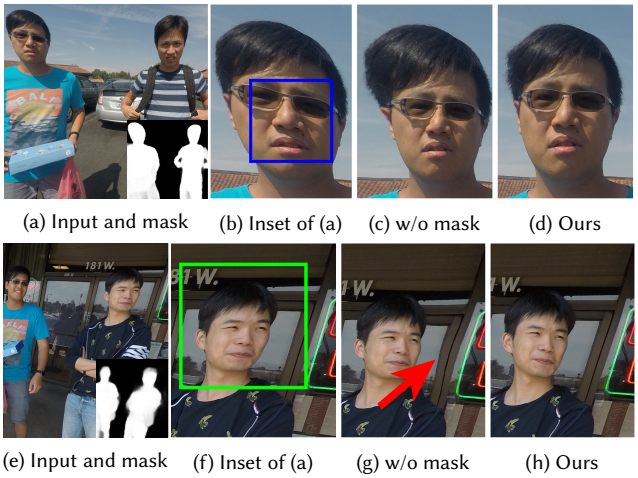
Face Objective Term. We introduce a novel per-face energy term, E_s, which has an one-to-one association with every face detected from the input image. The total face energy E_f is the sum of all the per-face energy terms:

$$E_f = \sum_k E_{s,k}, \tag{5}$$

where k indexes the detected faces. The per-face energy $E_{s,k}$ encourages facial regions to follow the stereographic mesh M_u by employing the formula below:

$$E_{s,k} = \sum_{i \in \mathbf{B}_k} w_i m_i \|\mathbf{v}_i - (\mathbf{S}_k \mathbf{u}_i + \mathbf{t}_k)\|_2^2 + \lambda(\mathbf{S}_k), \tag{6}$$

where $\{w_i\}$ are the face weights from Eq. (3), $\{\mathbf{u}_i\}$ are vertices on the stereographic mesh M_u, and $\mathbf{B}_k$ denotes the set of vertices on



(a) Input and mask (b) Inset of (a) (c) w/o mask (d) Ours

(e) Input and mask (f) Inset of (a) (g) w/o mask (h) Ours

Fig. 6. Subject masks are critical to our method (a)(e). **Top**: without using the subject mask and solely relying on a face bound (blue box in (b)), the result (c) misses hair regions and leads to artificial warping. **Bottom**: attempting to enlarge the face bound to cover the entire face (green box in (f)) introduces another artificial distortion at the boundary between the background and the face bounds (g), pointed by the arrow. Using subject masks, our method accurately identifies the regions to receive the correction (d)(h).



(a) Input (b) w/o Eq. (8) (c) w/o E_b (d) Ours

Fig. 7. Without the face scale term in Eq. (8), the input face in (a) is incorrectly magnified in the output (b), leading to the undesired look. Without bending terms in Eq. (9), rigid edges such as the window frames at the background are distorted, as pointed out by the arrow (c). With the combined energy terms, the result looks realistic and natural (d).



(a) Input (103° FOV) (b) Inset of (a) (c) No padding (d) Ours

Fig. 8. For faces close to the image border (a)(b), enforcing the hard boundary conditions causes artifacts at the image border, such as the elongated and unnatural lower chin in (c). We employ extra paddings to relax the boundary conditions, and output a more natural and realistic look (d).

the k-th face. Because the image corners have stronger perspective distortion and require more correction, we use $\{m_i\}$ sampled from a radial sigmoid function $(1 + \exp(-(r - r_a)/r_b))^{-1}$, where r is the radial distance of $\mathbf{p}_i$ on the input image , r_a and r_b are parameters controlling the attenuations of the correction strength, and chose such that m_i is 0.01 at the image center, and 1 at 100° FOV. Using the radial sigmoid function, our method leaves the image center unaffected and free from fish-eye artifacts at the center of stereographic projection. Intuitively, if we temporarily ignore other variables and concern only at $\sum_{i\in\mathbf{B}_k} w_i\|\mathbf{v}_i - \mathbf{u}_i\|_2^2$ in Eq. (6), the term favors the stereographic projection on the k-th facial region.

In Eq. (6), we propose a per-face similarity transform represented by $\mathbf{S}_k$ and $\mathbf{t}_k$, where $\mathbf{S}_k$ is parameterized by $[a_k, b_k]$:

$$\mathbf{S}_k = \begin{bmatrix} a_k & b_k \\ -b_k & a_k \end{bmatrix}, \tag{7}$$

and $\mathbf{t}_k$ is a two-dimensional translational vector. We note that $\mathbf{S}_k$ and $\mathbf{t}_k$ are hidden variables and will be discarded after the optimization. Their roles facilitate the optimizer to find a better vertices arrangement by slightly translating, rotating, and scaling each face individually. In case when a large group is present and faces are close to each other, the term reduces distorting artifacts caused by conflicting objectives, as demonstrated in Fig. 5.

The hidden variable a_k in Eq. (7) contributes to the face magnification factor in the output image. We add a regularization term to preserve the face scale:

$$\lambda(\mathbf{S}_k) = w_s\|a_k - s_t\|_2^2, \tag{8}$$

where w_s is the weight for the scale regularization term, and s_t is the face scaling target. We set w_s to 2000 and s_t to 1. Fig. 7b shows that Eq. (8) preserves consistent face sizes after the correction.

Using $\{w_i\}$, the face objective term takes the boundary between the face and background into account. The benefits are clearly shown in Fig. 6.

Line-Bending Term. On the boundary between the face and background, straight lines may be distorted because the two regions follow different projections. We preserve straight lines by encouraging the output mesh to scale rather than twist by adding a line-bending term [Chang et al. 2011]:

$$E_b = \sum_i \sum_{j\in N(i)} \|(\mathbf{v}_i - \mathbf{v}_j) \times \mathbf{e}_{ij}\|_2^2, \tag{9}$$

where $\mathbf{e}_{ij}$ is the unit vector along the direction $\mathbf{p}_i - \mathbf{p}_j$, and $\times$ denotes the cross product. The line-bending term penalizes the shearing of the grid, and therefore preserves the edge structure on the background. Figs. 7c and 7d show the comparison with and without this term.

Regularization Term. Finally, we regularize the mesh by encouraging smoothness between 4-way adjacent vertices using a regularization term E_r:

$$E_r = \sum_i \sum_{j\in N(i)} \|\mathbf{v}_i - \mathbf{v}_j\|_2^2. \tag{10}$$

Discussion. We have tested local perspective projection with the projection center in the middle of the face, but this method fails when multiple faces clutter together. Since these local projections are conflicting with each other, we observe strong distortions along the face boundary. Instead, the stereographic projection is a global

projection and reconciles faces at various locations. We have experimented the grid line-bending regularization terms by Wang et al. [2008]. It works effectively but introduces a non-linear optimization, which requires alternative minimization and becomes challenging for interactive usage on mobile platforms.

4.1 Mesh Boundary Extension

The energy function combining Eqs. (6), (9) and (10) has a trivial null solution ($\mathbf{v}_i = 0\ \forall i$). A simple boundary condition by forcing $\mathbf{v}_i = \mathbf{p}_i$ on the mesh boundary would avoid the problem. However, it creates strong distortions when faces are close to image boundary, as shown in Fig. 8c.

We expand the mesh domain by padding q vertices on each side. We then enforce the following boundary conditions *on the extended mesh*:

$$\begin{cases} v_{i,x} = p_{i,x}, & \text{if } i \in \text{left or right boundary,} \\ v_{i,y} = p_{i,y}, & \text{if } i \in \text{top or bottom boundary.} \end{cases} \tag{11}$$

That is, the boundary vertices can only move along borders. We set q to 4. The padded vertices will be cropped out after the optimization. The domain extension in Eq. (11) distributes the distortion to padded vertices, and reduces artifacts near the boundary of the output image, as shown in Fig. 8.

To reduce undefined regions after mesh warping, we encourage the vertices *on the original mesh boundary* to expand instead of shrinking using the asymmetric cost term E_a:

$$E_a = E_l + E_r + E_t + E_b\,, \tag{12}$$

$$\begin{aligned} E_l &= \mathbb{I}(v_{i,x} > 0) \cdot \|v_{i,x}\|_2^2, \forall i \in \partial_{\text{left}}\,, \\ E_r &= \mathbb{I}(v_{i,x} < W) \cdot \|v_{i,x} - W\|_2^2, \forall i \in \partial_{\text{right}}\,, \\ E_t &= \mathbb{I}(v_{i,y} > 0) \cdot \|v_{i,y}\|_2^2, \forall i \in \partial_{\text{top}}\,, \\ E_b &= \mathbb{I}(v_{i,y} < H) \cdot \|v_{i,y} - H\|_2^2, \forall i \in \partial_{\text{bottom}}\,, \end{aligned} \tag{13}$$

where $\mathbb{I}(.)$ is the indicator function that returns 1 for true condition and 0 otherwise, W and H are image dimensions, and ∂_* are the original mesh boundary. The benefits of E_a are more noticeable when faces are close to image border.

4.2 Optimization

The total energy function is the weighted combination of the energies from Eqs. (6), (9), (10) and (12):

$$\begin{aligned} E_t = \lambda_f E_f + \lambda_b E_b + \lambda_r E_r + \lambda_a E_a\,, \\ \text{subject to Eq. (11),} \end{aligned} \tag{14}$$

where λ_f, λ_b, λ_r, and λ_a are the weights for the corresponding energy terms, and fixed to 4, 2, 0.5, and 4, respectively, for all the results in this work.

To accelerate the optimization process, we initialize vertices $\{\mathbf{v}_i^0\}$ by interpolating perspective and stereographic mesh, $\{\mathbf{p}_i\}$ and $\{\mathbf{u}_i\}$, with weights $\{w_i\}$ from Eq. (3) using the kernel regression:

$$\begin{aligned} \mathbf{v}_i^0 &= \mathbf{p}_i + \Delta\mathbf{v}_i^0\,, \\ \Delta\mathbf{v}_i^0 &= \frac{\sum_j w_j K_h(\|\mathbf{p}_j - \mathbf{p}_i\|_2^2)\Delta\mathbf{v}_j}{\sum_j w_j K_h(\|\mathbf{p}_j - \mathbf{p}_i\|_2^2)}\,, \\ \Delta\mathbf{v}_j &= \begin{cases} \mathbf{u}_j - \mathbf{p}_j & \text{if } \mathbf{p}_j \in \{\mathbf{B}_k\}\,, \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \tag{15}$$

where $K_h(x) = \exp(-x/(2h^2))$ is the Gaussian interpolation kernel with bandwidth parameter h set to 2.37. Our initialization favors stereographic projection over facial regions, and brings the grid closer to the final solution and speeds up the convergence. After the optimization process and removing the padded vertices, we shift and zoom the output to avoid undefined regions after warping.

To frame the warped output, we linearly normalize M^* to $M_n = \{\mathbf{v}_{n,i}\}$ with a scale s_g and translation $\mathbf{t}_g$:

$$\mathbf{v}_{n,i} = s_g(\mathbf{v}_i^* + \mathbf{t}_g). \tag{16}$$

We set $\mathbf{t}_g = -\mathbf{v}_0^*$ to map the top-left mesh corner to the origin. The global scale is retrieved by traversing the boundary vertices to search for the minimal zoom factor.

4.3 Implementation Details

We set the mesh dimension to 103×78 for all of our results. We implement our solution on both the desktop and the Qualcomm Snapdragon 845 mobile platform (SDM845) . The mobile face segmenter runs efficiently on Adreno 630 GPU using TensorFlow Lite [TensorFlow 2018]. We solve the constrained least-squared problem in Eq. (4) with the Ceres solver [Agarwal et al. 2012], and apply the Trust Region minimizer [Bertsekas 1999] with a sufficiently large initial trust region. To warp the input from the perspective to the optimized mesh, *i.e.*, M_p to M_n, we use a high-quality Lanczos resampler [Duchon 1979] with 7×7 support in a GPU shader.

The memory consumption peaks at image warping, which requires an additional full-sized frame buffer. The total running time for a 12 mega-pixel input takes 920ms on SDM845, with a major breakdown to segmentation, mesh optimization, and warping at 280ms, 340ms, and 115ms, respectively.

All the results in this paper and the supplemental materials are generated from the same parameter setting. We validate our parameters on a small dataset of 20 photos containing single selfies and group photos and apply them for all results in the paper. We develop an interactive desktop user interface to facilitate the task. We sequentially determine the weights in Eq. (14) for face term, regularization term, bending term, and boundary term. Since these terms address artifacts orthogonal to each other, we achieve good results after 2-3 iterations.

Our method does not include calibration. The only required input is the camera FOV, which is required by Eq. (1). We determine the FOV by retrieving the focal length and the sensor information from EXIF metadata embedded in the image. Additionally, we correct the

lens distortion for the input image using the embedded distortion coefficients.

For performance optimization, we combine the warping for the lens distortion correction and the proposed local undistortion into a single step, as illustrated in the supplemental material. Throughout this paper, we show the lens distortion corrected images as inputs. Due to rectangular crop after lens correction, the input may show less content than the output near the boundary.

5 RESULTS

Fig. 9 shows our perspective distortion correction for wide-angle portraits taken from cameras of various FOVs. Our inputs cover from single-person photos to group of five. In these results, our method successfully corrects perspective distortion on all faces, and renders natural and realistic looks. In challenging backgrounds containing human-made objects such as buildings, windows, doors, furnitures, and interior structures, our method preserves their geometries without introducing artifacts. Our method performs well in different lighting conditions, including low-light and harsh sunlight. We show the optimized mesh of each result in the supplemental material.

To verify the robustness of our method, we evaluate it on 167 photos across different countries, group sizes, camera models, and various FOVs ranging from 70° to 120°, collected by us and from Flickr. The Flickr images are collected by searching for keywords "wide-angle" and "people". The results are shown in the supplemental material. The dataset contains a large variety of casual wide-angle photos with various facial features such as glasses, hair styles, accessories, as well as people of different genders, ages, and facial expressions. It also includes a variety of uncontrolled lighting conditions, both indoors and outdoors, in front of landscape and cityscape background full of human-made objects . This dataset is challenging because some photos are noisy due to low-light conditions, and the background can be cluttered, which makes precise segmentation difficult.

Correction at Various FOVs. Fig. 10 shows the images of a controlled experiment. The face positions are spanned over the entire view frustum of a 97° FOV camera. Fig. 10a clearly shows how the perspective projection distorts facial features depending on the subject location. Our correction in Fig. 10b makes faces across camera FOV look consistent and similar to the face shot at the camera center. Note that the face at image center is nearly identical before and after the correction.

Normal FOVs. Perspective distortion also appears on the normal FOV cameras. Fig. 11 shows that our correction makes noticeable improvements on a 78° FOV camera, which is roughly the specification of the major premium phones in our study[1]. Fig. 12 shows a typical family portrait with 20 subjects. The faces near the corners are stretched and distorted. Our method seamlessly corrects those faces without affecting the overall geometry of the scene.

[1]Apple iPhoneX (front/rear): 78° FOV/75.4° FOV, Samsung Galaxy S9: 80° FOV/77° FOV, Huawei Mate 20 Pro: 79.6° FOV/77.4° FOV.

Conformality Cost Reduction. Carroll et al. [2009] defines conformality cost as a measurement quantifying the amount of perspective distortion on a photograph. We show the reduction of conformality cost using their formula in Fig. 13. To account for an arbitrary warping mesh in conformality cost calculation, we adopt the prior work [Carroll et al. 2009; Zorin and Barr 1995] to propagate the conformality using the chain rule in calculus to compute the Jacobian. After our correction, the conformality costs on facial area are significantly reduced, and evenly distributed to the non-facial regions.

Bokeh Mode Enhancement. Fig. 14 shows that our method improves the image quality for Bokeh effect. As the in-focus human face is the key subject now, the perspective distortion becomes even more unflattering. Consequently, Bokeh mode on mobile phones often have to reduce FOV by cropping zoom [Wadhwa et al. 2018]. Our method resolves the issue and improves the image quality of Bokeh mode on wide-angle cameras.

Challenging Cases. Fig. 15 shows results of challenging face poses taken from side views, substantial occlusions by sunglasses, a hat, covered and wrapped by arms. The tests are taken at difficult lighting conditions including back-lit and low-light scenarios. Our method successfully corrects faces in these cases without introducing artifacts.

5.1 Comparison with Related Work

Figs. 2 and 3c provide the comparisons against Mercator and the stereographic projections commonly used in panorama photography. Our method locally corrects face regions without affecting the background, while other methods bend the entire scene.

Fig. 16 shows the results by our algorithm and the state-of-the-art Pannini projection method [Sharpless et al. 2010]. The Pannini projection fails to fully correct the distortion for the subject at the corner. Further, it falsely magnifies the face at the image center. This is a challenging case because the subject at corner requires strong correction, while the center one should not be affected at all. Our method applies local warping to make the subject at the corner look natural, and also maintains the look for the subject at the center.

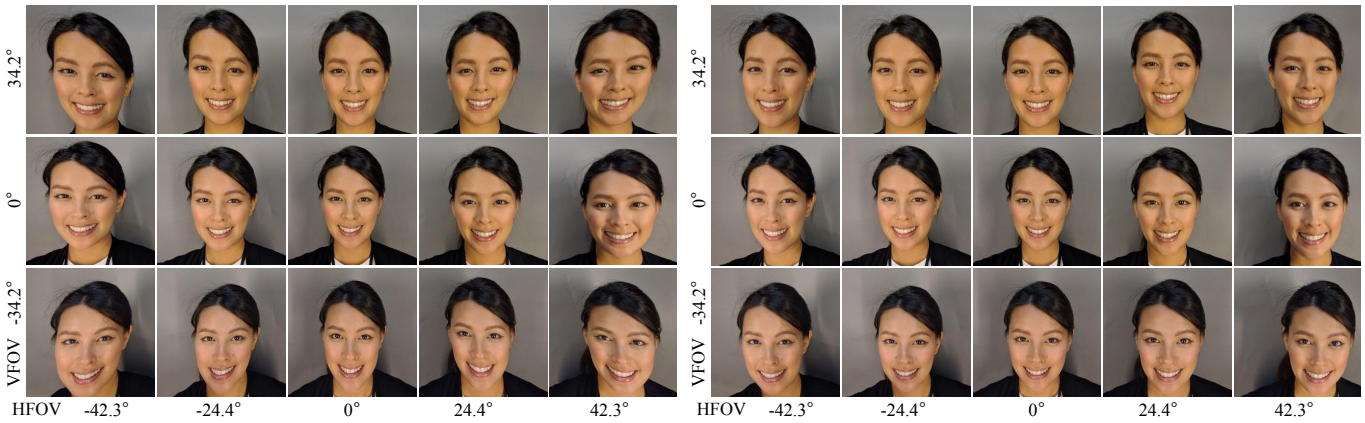
Fig. 17 shows the results by our method and the method by Zorin and Barr [1995], which corrects geometric perceptual distortions using a parametric warping. We use the inputs and their results from the original paper. While both methods successfully restore the geometric distortion on the subject's face, their method distorts the coffee table and the wall at the right-hand side in the example at the top row, and the monitors at the bottom row. With the help of the locally optimized warping, our method retains the shape of these objects.

5.2 Comparison with Commercial Products

We compare our method with those by commercial products for distortion correction and perspective manipulations.

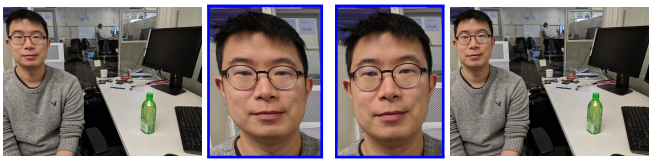


Fig. 9. Each pair shows an input distorted by the perspective projection (left), and our result (right). We include common indoor and outdoor scenes with various numbers of subjects and natural/human-made backgrounds. Our method recovers all face shapes without artifacts. ©Tony Nungaray



(a) Input: perspective projection. (b) Our results at the corresponding positions.

Fig. 10. In a controlled experiment, the subject is positioned at 15 locations over a uniform 5×3 grid within the view of a diagonal 97° FOV camera, labeled by the horizontal and vertical FOV (HFOV and VFOV, respectively) on the two axes. We crop the faces to composite (a) to highlight the perspective distortion at different positions, especially at corners and edges. We then process each full-size image in (a), and crop the face on the output to composite (b). Our method successfully recovers face shapes at any positions and generates natural looks.



(a) Normal-angle 78° FOV input (b) Our method

Fig. 11. Current major premium phones ship main cameras around 78° FOV[1]. The perspective distortion on this FOV is clearly visible (a). Our method corrects the distortions on faces with noticeable quality improvements (b).



Fig. 12. **Left**: a group shot with 20 subjects taken by a 79° FOV camera. **Right**: the results of our correction. The insets in the bottom row highlight that our method performs well for a large group. ©Daniel Kleeman



(a) Input (97° FOV) (b) Our method

Fig. 13. **Top:** an image pair before and after our correction. **Bottom:** we overlay the top row with the conformality cost map [Carroll et al. 2009]. Areas with higher perspective distortion are labeled with warmer colors. For an input rendered with the perspective projection, the conformality cost isomorphically increases from the image center to the corners (a). Our method restores the distorted faces and reduces the conformality cost on the face regions (b).

Adobe Photoshop "Perspective Warp". Fig. 16c shows the image after manual perspective correction using Adobe Photoshop Perspective Warp feature [Intwala and Agarwala 2015], which changes the perspective of a photograph by manually adjusting the control points. When trying to correct the distortion for the subject at the corner, the perspective warp inevitably creates another distortion to the center subject, and leads to irregular image boundary. Our method is automatic and performs well for faces at any positions.

DxO ViewPoint "Volume Deformation". Fig. 18 shows the results by our algorithm and the Volume Deformation Correction feature in DxO ViewPoint [DxO 2018]. To render the final results, users are required to manually adjust the strength of the correction, while our method is fully automatic. Similarly to global projections, DxO's solution distorts the straight lines at the output, *e.g.* the fences and the facades behind the subjects, and loses image content nearby



(a) Bokeh w/o our method (103° FOV) (b) With our method

Fig. 14. Bokeh (synthetic defocus) effect draws user's attentions to subjects' faces. (a) Perspective distortion is particularly unpleasing as it skews the face. (b) Our algorithm corrects the distortion and creates the output with the pleasant look.

image border. Our method applies correction locally to preserve straight lines in the scene, and keeps the output FOV close to the input.

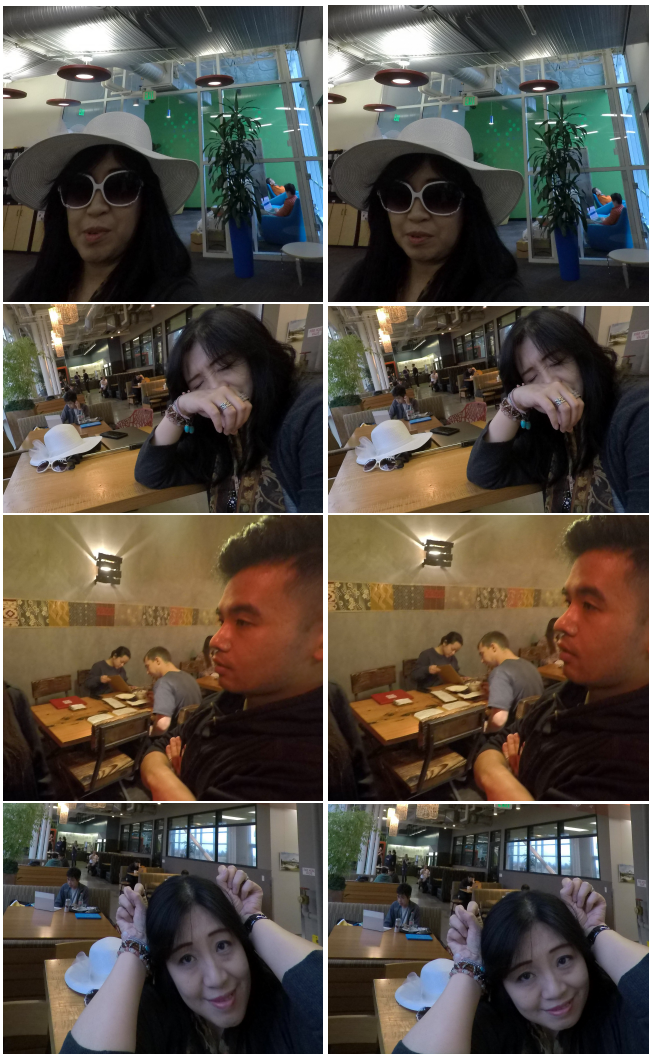
Samsung Galaxy "Face Shape Correction". Fig. 19 shows the results by our method and Samsung Galaxy S9+ "Face Shape Correction" feature, which shares with our goal to automatically correct wide-angle distortion [Samsung 2017]. Both methods fix perspective distortion for faces at corners. However, S9+ introduces artifacts that make faces at the center appear larger than that without the correction, and distorts the straight lines on the window frames in the background. Our method generates natural and realistic looks for all the faces and background.

5.3 User Study

We evaluate our method on 4131 photos collected by us and from Flickr, ranging from 70° to 120° FOV, containing from 1 to 20 faces. We randomly sample 1047 photos from the 4k+ images for user study performed on Amazon Mechanic Turk. For each image, we show both perspective projection and the result by our method with randomized left-right order, and ask 5 testers to select the one that "looks more natural with less distortion." We conclude the favored one by major voting from the 5 testers to reduce noises. In this study, 92.4% of our results are classified as more natural. For reference, when compared with the stereographic, Mercator, and Pannini projections [Sharpless et al. 2010], the ratios favoring our method are 84%, 81%, and 85%, respectively. We describe the details in the supplemental document.

6 LIMITATIONS AND CONCLUSION

Our approach requires face detection and segmentation. As demonstrated in Fig. 20, missing face detection or erroneous subject segmentation fails in correction or introduces artifacts. The curvature of an edge may change when it is very close to face regions, as shown in Fig. 21. Our method is limited to face regions. As illustrated in Fig. 22, correcting faces without the rest of body, such as torso, shoulders, or arms, may create artificial looks due to the inconsistency, while most users still prefer faces without distortions. Our method does not perform with faces printed on a poster, since



(a) Input (103° FOV) (b) Our method

Fig. 15. (a) Inputs at challenging cases. From top to bottom row: accessories that cover eyes and hairs, gestures occluding faces while the subject closes the eyes, profile portraits taken at low light conditions, and arms around faces. (b) Our method performs well on these challenging cases.

the pinhole camera model does not apply. Without knowing the camera model for the poster, our algorithm can over or under correct it. Similarly, distorted architectures sometimes look artificial nearby corrected faces. We expect to address the issue with advanced object segmentation technique. We do not correct distortions caused by foreshortening, which appear when faces are very close to the camera. We review the 4131 photos mentioned in Sec. 5.3. Our method fails in 7 cases (0.16%) due to missing segments on small subjects at the background (Fig. 20), and 6 cases (0.1%) due to distortions on objects nearby the face (Fig. 21).

Conclusion and Future Work. We have proposed an automatic algorithm to correct perspective distortion on wide-angle portrait



(a) Input (100° FOV) (b) Pannini [Sharpless et al. 2010] (c) Adobe Photoshop Perspective Warp (d) Our method

Fig. 16. We evaluate our method with the automatic correction using (b) Pannini warp, which unfortunately under-corrects the face at the top-left corner. (c) We manually edit (a) using Perspective Warp in Photoshop [Intwala and Agarwala 2015] to correct the face at the top-left corner. The manipulation however changes the camera perspective, and introduces another distortion on the subject at the image center. (d) Our method recovers the face shape of the subject at the top-left corner without alternating the rest of the input.



(a) Input (b) Zorin and Barr (c) Our method

Fig. 17. We evaluate our algorithm against the manual method by Zorin and Barr [1995] at the top row, and their automatic method at the bottom row. Both approaches can recover the face shapes. However, their method distorts the human-made objects such as the coffee table, the walls, and the monitors, while our algorithm does not affect them. The input FOV for the top and bottom rows are 100° and 92°, respectively. ©Denis Zorin and ©Dave Dooling.



(a) Input (97° FOV) (b) DxO ViewPoint3 (c) Our method

Fig. 18. We evaluate our method against the commercial product, DxO ViewPoint3 "Volume Deformation Correction" feature, which requires manual editing to address perspective distortion. DxO ViewPoint3 unfortunately distorts the fences and buildings in the background, and loses the image content at the corners. Our method performs well by preserving their shape and FOV from the input.

photos. Using an energy minimization, our local face undistortion combines the stereographic and perspective projections onto a single image, and adapts to face regions. We tested the proposed method on a wide variety of photos with different FOVs and scenes to show its robustness. We believe this work encourages users enjoying wide-angle photography.



(a) Input (80° FOV) (b) Samsung Galaxy S9+ (c) Our method

Fig. 19. Evaluation against Samsung Galaxy S9+ "Face Shape Correction" (FSC) feature. (a) A photo captured by S9+ with FSC disabled. The face at the top-left corner exhibits perspective distortion. (b) FSC corrects the face but introduces new artifacts: the face at the center appears larger, and the window frames are distorted. (c) Our method makes all the faces and background distortion-free.



(a) Input (104° FOV) (b) Our method

Fig. 20. A face detector may miss some subjects in crowded scenes (a) due to tiny face sizes. While our method properly corrects the detected face (green box), it fails for the undetected face (red box). ©Tony Nungaray

We are interested in extending the correction to wide-angle videos. We expect new challenges in temporal coherence and real-time requirement on mesh optimization. For example, subjects may appear and disappear during the video recording, and user zoom may continuously changes the camera FOV. Another extension is to correct perspective distortions on buildings and salient objects. Beyond the graphics application, perhaps a deeper question is this: *under perspective projection, what image features cause human vision system to feel distorted, and how to address them?* We are excited to see more research on wide-angle images.



(a) Input (97° FOV) (b) Our method

Fig. 21. In a case containing multiple faces (a), the straight edge close to the subject at the top-left corner is unfortunately bent in our output (b).



(a) Input (90° FOV) (b) Our method

Fig. 22. Our method restores the face but not torso for the left-most subject, and creates an inconsistent look. ©Tony Nungaray

ACKNOWLEDGMENTS

We thank the reviewers for numerous suggestions on user study and exposition. We also thank valuable inputs from Ming-Hsuan Yang, Marc Levoy, Timothy Knight, Fuhao Shi, and Robert Carroll. We thank Yael Pritch, David Jacobs, Neal Wadhwa, Juhyun Lee, and Alan Yang for supports on subject segmenter and face detector integration, Kevin Chen and Sung-fang Tsai for GPU acceleration. We thank Sam Hasinoff, Eino-Ville Talvala, Gabriel Nava, Wei Hong, Lawrence Huang, Chien-Yu Chen, Zhijun He, Paul Rohde, Ian Atkinson, and Jimi Chen for supports on mobile platform implementations. We thank Weber Tang, Jill Hsu, Bob Hung, Kevin Lien, Joy Hsu, Blade Chiu, Charlie Wang, and Joy Tsai for image quality feedbacks, Karl Rasche and Rahul Garg for proofreading. Finally, we give special thanks to Denis Zorin and all the photography models in this work for photo usage permissions and supports on data collection.

REFERENCES

Michael D Abràmoff, Paulo J Magalhães, and Sunanda J Ram. 2004. Image processing with ImageJ. *Biophotonics international* 11, 7 (2004), 36–42.
Sameer Agarwal, Keir Mierle, and Others. 2012. Ceres Solver. http://ceres-solver.org.
Aseem Agarwala, Maneesh Agrawala, Michael Cohen, David Salesin, and Richard Szeliski. 2006. Photographing long scenes with multi-viewpoint panoramas. In *ACM TOG*, Vol. 25. 853–861.
Dimitri P Bertsekas. 1999. *Nonlinear programming*.
Duane C Brown. 1966. Decentering distortion of lenses. *Photogrammetric Engineering and Remote Sensing* (1966).
Robert Carroll, Aseem Agarwala, and Maneesh Agrawala. 2010. Image warps for artistic perspective manipulation. In *ACM TOG*, Vol. 29. 127.
Robert Carroll, Maneesh Agrawal, and Aseem Agarwala. 2009. Optimizing content-preserving projections for wide-angle images. In *ACM TOG*, Vol. 28. 43.
Che-Han Chang and Yung-Yu Chuang. 2012. A line-structure-preserving approach to image resizing. In *Proc. CVPR*. 1075–1082.
Che-Han Chang, Chia-Kai Liang, and Yung-Yu Chuang. 2011. Content-aware display adaptation and interactive editing for stereoscopic images. *IEEE TMM* 13, 4 (2011), 589–601.
Che-Han Chang, Yoichi Sato, and Yung-Yu Chuang. 2014. Shape-preserving half-projective warps for image stitching. In *Proc. CVPR*. 3254–3261.
Yu-Sheng Chen and Yung-Yu Chuang. 2016. Natural image stitching with the global similarity prior. In *Proc. ECCV*. 186–201.
Song-Pei Du, Shi-Min Hu, and Ralph R Martin. 2013. Changing perspective in stereoscopic images. *IEEE TVCG* 19, 8 (2013), 1288–1297.
Claude E Duchon. 1979. Lanczos filtering in one and two dimensions. *Journal of applied meteorology* 18, 8 (1979), 1016–1022.
DxO. 2018. Correcting volume deformation with DxO ViewPoint. https://www.dxo.com/dxo-viewpoint/features/.
Ohad Fried, Eli Shechtman, Dan B Goldman, and Adam Finkelstein. 2016. Perspective-aware manipulation of portrait photos. *ACM TOG* 35, 4 (2016), 128.
Ran Gal, Olga Sorkine, and Daniel Cohen-Or. 2006. Feature-aware texturing. *Rendering Techniques* 2006, 17 (2006), 2.
GoPro. 2017. HERO6 Black Field of View (FOV) Information. https://gopro.com/help/articles/question_answer/hero6-black-field-of-view-fov-information.
Heng Guo, Shuaicheng Liu, Tong He, Shuyuan Zhu, Bing Zeng, and Moncef Gabbouj. 2016. Joint video stitching and stabilization from moving cameras. *IEEE TIP* 25, 11, 5491–5503.
Kaiming He, Huiwen Chang, and Jian Sun. 2013. Rectangling panoramic images via warping. *ACM TOG* 32, 4 (2013), 79.
Chintan Intwala and Aseem Agarwala. 2015. Perspective warp. US Patent 9,117,253.
Hamid Izadinia, Bryan C Russell, Ali Farhadi, Matthew D Hoffman, and Aaron Hertzmann. 2015. Deep classifiers from image tags in the wild. In *Proc. 2015 Workshop on Community-Organized Multimodal Mining: Opportunities for Novel Solutions*. 13–18.
Yoshihiro Kanamori, Nguyen Huu Cuong, and Tomoyuki Nishita. 2011. Local optimization of distortions in wide-angle images using moving least-squares. In *Proceedings of the 27th Spring Conference on Computer Graphics*. ACM, 51–56.
Johannes Kopf, Dani Lischinski, Oliver Deussen, Daniel Cohen-Or, and Michael Cohen. 2009. Locally adapted projections to reduce panorama distortions. In *CGF*, Vol. 28. 1083–1089.
Hyunjoon Lee, Eli Shechtman, Jue Wang, and Seungyong Lee. 2012. Automatic upright adjustment of photographs. In *Proc. CVPR*. 877–884.
Tommer Leyvand, Daniel Cohen-Or, Gideon Dror, and Dani Lischinski. 2008. Data-driven enhancement of facial attractiveness. *ACM TOG* 27, 3 (2008), 38.
Dongping Li, Kaiming He, Jian Sun, and Kun Zhou. 2015. A geodesic-preserving method for image warping. In *Proc. CVPR*. 213–221.
Feng Liu, Michael Gleicher, Hailin Jin, and Aseem Agarwala. 2009. Content-preserving warps for 3D video stabilization. In *ACM TOG*, Vol. 28. ACM, 44.
Feng Liu, Michael Gleicher, Jue Wang, Hailin Jin, and Aseem Agarwala. 2011. Subspace video stabilization. *ACM TOG* 30, 1 (2011), 4.
Darko Pavić, Volker Schönefeld, and Leif Kobbelt. 2006. Interactive image completion with perspective correction. *The Visual Computer* 22, 9-11 (2006), 671–681.
Samsung. 2017. Galaxy S7: About Shape Correction. www.samsung.com/nz/support/mobile-devices/galaxy-s7-about-shape-correction/.
Thomas K Sharpless, Bruno Postle, and Daniel M German. 2010. Pannini: a new projection for rendering wide angle perspective images. In *Proc. 6th international conference on Computational Aesthetics in Graphics, Visualization and Imaging*. 9–16.
John Parr Snyder. 1987. *Map projections–A working manual*. Vol. 1395. US Government Printing Office.
Mahdi Abbaspour Tehrani, Aditi Majumder, and M Gopi. 2016. Correcting perceived perspective distortions using object specific planar transformations. In *Proc. ICCP*.
TensorFlow. 2018. TensorFlow-Mobile/Lite. https://www.tensorflow.org/mobile/.
Kim H Veltman. 1986. Perspective, anamorphosis and vision. *Marburger Jahrbuch für* (1986).
Dhanraj Vishwanath, Ahna R Girshick, and Martin S Banks. 2005. Why pictures look right when viewed from the wrong place. *Nature neuroscience* 8, 10 (2005), 1401.
Neal Wadhwa, Rahul Garg, David Jacobs, Bryan Feldman, Nori Kanazawa, Robert Carroll, Yair Movshovitz-Attias, Jonathan Barron, Yael Pritch, and Marc Levoy. 2018. Synthetic depth-of-field with a single-camera mobile phone. *ACM TOG* 37, 4 (2018).
Yu-Shuen Wang, Chiew-Lan Tai, Olga Sorkine, and Tong-Yee Lee. 2008. Optimized scale-and-stretch for image resizing. In *ACM TOG*, Vol. 27. 118.
Jin Wei, Chen-Feng Li, Shi-Min Hu, Ralph R Martin, and Chiew-Lan Tai. 2012. Fisheye video correction. *IEEE TVCG* 18, 10 (2012), 1771–1783.
Lior Wolf, Moshe Guttmann, and Daniel Cohen-Or. 2007. Non-homogeneous content-driven video-retargeting. (2007).
Lihi Zelnik-Manor, Gabriele Peters, and Pietro Perona. 2005. Squaring the circle in panoramas. In *Proc. ICCV*, Vol. 2. 1292–1299.
David Zhang, Qijun Zhao, and Fangmei Chen. 2011. Quantitative analysis of human facial beauty using geometric features. *Pattern Recognition* 44, 4 (2011), 940–950.
Fang-Lue Zhang, Xian Wu, Hao-Tian Zhang, Jue Wang, and Shi-Min Hu. 2016. Robust background identification for dynamic video editing. *ACM TOG* 35, 6 (2016), 197.
Zhengyou Zhang. 2000. A flexible new technique for camera calibration. *IEEE Trans. PAMI* 22 (2000).
Denis Zorin and Alan H Barr. 1995. Correction of geometric perceptual distortions in pictures. In *SIGGRAPH*. 257–264.