Adaptive Content-aware Correction for Wide-angle Portrait Photos

Juan Cao, Binyan Lin
School of Mathematical Sciences
Xiamen University
juancao@xmu.edu.cn

Zhonggui Chen
School of Informatics
Xiamen University
chenzhonggui@xmu.edu.cn

Abstract

Portraits near the periphery of wide-angle photos often suffer conspicuous distortion. With the popularity of wide-angle lenses on mobile phones, portrait correction, which removes the portrait distortion of wide-angle images, has been attracting widespread attention in the field of content-aware warping. Existing portrait correction methods for wide-angle photos using uniform quad meshes take a long time solving the correction optimization, and most of them focus only on correcting faces, leading to the inconsistency of heads and bodies after correction. In this paper, we propose an efficient method based on triangle mesh to remove portrait distortion in wide-angle perspective photos. We generate an adaptive mesh tailored to the image content with fewer vertices. Relying on the characteristics of the triangle mesh, we tailor three smooth and intuitive energy terms for the human area, background area, and boundary to minimize portrait distortion. Our algorithm has good expandability in adding more geometric constraints, such as line constraint. Experimental results show that our method is robust for photos with various fields of view. Comparisons with the state-of-the-art demonstrate that our method achieves significant improvements in optimization efficiency and consistency of heads and bodies.

1. Introduction

People often record interesting moments by taking photos with phones that have a wide field of view (FOV). A natural look can be achieved with a narrow lens FOV, but with the increase of the lens' view angle, the distortion becomes stronger the closer one is to the image boundary [30]. Zorin *et al*. proved that single global projection cannot balance the shapes of salient objects and the straightness of lines [37]. Doing so would produce serious stretches in perspective images and apparent bending lines in fish-eye images due to their various viewing changes. Photos with excessive stretch and bending line structure are not consistent with our subjective perception. Therefore, the correction of wide-angle photos is necessary.

Many wide-angle portrait distortion methods have been proposed, such as parametric methods [37, 22], least square methods [6, 29, 27] and deep learning method [28]. However, most of them focus only on face correction and do not take the body into consideration [6, 27, 28, 22]. Some works need to specify the portrait segmentation by users [29]. In addition, most wide-angle portrait correction algorithms use uniform quad mesh, where the vertex distribution in the non-salient area is as dense as the salient area, leading to time inefficiency in minimizing distortion.

In this paper, we present an adaptive content-aware correction approach to correct portrait distortion in wide-angle perspective photos. We formulate the problem into a nonlinear least square optimization, which measures the distortion based on the adaptive triangle mesh. We adapt stereographic projection to the portrait regions that contain faces and bodies to ensure the consistency of the head-to-body ratio in the result. We also preserve the perspective projection for background regions. User-specified geometric constraints, such as line constraint, can be easily incorporated into our optimization framework to satisfy users' special preferences.

We experimentally demonstrate that our algorithm can produce high-quality correction results. In addition, our method significantly improves the effectiveness of the optimization process due to the reasonable quantity and distribution of mesh vertices. Specifically, we make the following contributions in this paper:

- We adopt the adaptive triangle mesh in wide-angle portrait correction. Therefore, we can drive the correction using fewer vertices to improve the efficiency of the correction optimization.
- We tailor the smooth and efficient energy function for the adaptive mesh. Users-specified geometry constraints, *e.g*., line constraint, can be easily incorporated into the energy function.
- We design an appropriate density map to optimize the adaptive mesh, improving the mesh quality by adjusting the vertex distribution such that the adaptive mesh fit the portraits well.

This paper is organized as follows. In Section 2 , we discuss related wide-angle correction works. In Section 3, we describe the portrait correction problem and overview of our algorithm. In Section 4, we generate the adaptive triangle mesh based on the density map generated by the portrait mask. In Section 5, we construct the optimization problem which measures the distortion using three content-aware energy terms. In Section 6, we show our results and comparisons with the state-of-the-art. The conclusion and future works are shown in Section 7.

2. Related Work

The core issue of this paper is to correct the distortion of portraits in wide-angle photos. We discuss the most relevant ones to our work: global distortion manipulation and content-aware portrait correction.

2.1. Global distortion manipulation

Wide-angle photos commonly suffer from conspicuous distortion due to inherent distortion and specific projection models. Although many calibration methods have been proposed [36, 3, 24, 16, 14, 2, 4] to remove the inherent distortion, they mainly focus on correcting straight lines to obtain perspective projection images. When the camera's FOV increases, the objects near the periphery of perspective projection photos will stretch significantly. Some works adjust the image perspective by changing the camera's pose [17, 9] or constraining vanish points and lines [5, 25]. However, these methods mainly reduce the perspective of architecture and landscape pictures instead of portrait photos. Stereographic [32], Mercator, and Pannini projections [26] are usually used to alleviate perspective distortion in portrait regions, but they bend straight lines in the background. The above-mentioned works are all global distortion manipulation that applies the same transformation for the different contents in an image. Therefore, they are unsuitable for moving the distortion of portraits while preserving straight lines in the background.

2.2. Content-aware portrait correction

Portrait correction in wide-angle photos is a special content-aware warping that is widely used in image resizing [15, 7, 8, 31], panorama stitching [13, 19], texture mapping [11], and video stabilization [10, 35]. Portrait correction methods have two categories: mesh-based portrait corrections and learning-based algorithms.

Mesh-based portrait correction. Zorin *et al*. proved that it's impossible to construct a viewing transformation such that all lines are straight and all spheres are exact circles in an image [37]. Therefore, Zorin *et al*. presented a parametric approach that combines the perceptual characteristics of plane and sphere projection to alleviate the perspective distortion. However, their work does not separate the portrait and background regions during the deformation. Nguyen *et al*. established an adaptive polynomial model, taking the faces into account independently [22]. However, Nguyen's model cannot dynamically adjust the vertex distribution of the transition area between the portrait and background area.

To optimize the vertex distribution in the transition area, many works formulated the correction problem as a least squares optimization by minimizing the energy terms that measure wide-angle image distortion. Carroll *et al*. used the conformality invariance of cylindrical projection and straight-line constraints specified by users to drive the vertices of the uniform quad mesh to undistort the salient area [6]. Shih *et al*. presented an automatic content-aware algorithm that adopts stereographic projection in face regions and perspective projection in other regions [27]. Tehrani *et al*. applied object-specific planer transformation on segmented image plane to minimize distortion [29].

All the above methods drive the correction by using uniform mesh. The mesh vertices in the non-salient area are as dense as the salient area, which is why the least squares problem would have too many variables, thereby leading to inefficient performance in the optimization process [29, 6, 27]. Moreover, these methods mainly focus on face correction, which often produces results with inconsistent head-to-body ratios.

Learning-based algorithm. Deep learning approaches for portrait correction have also been proposed, but most of them focus on the improvement of facial attractiveness [18], such as face normalization [21], shapely portraits [34], double chin removal [33]. Tan *et al*. first proposed a deep learning algorithm to correct the portraits for wide-angle photos [28]. Their work can be viewed as the generalization of Shih's work [27] in the deep learning field. Although the learning method can correct the distortion automatically without parameters, it needs many training images for learning.

Therefore, a more efficient method of generating corrective images is necessary. Inspired by previous works, we apply a portrait correction algorithm that is driven by adaptive triangle mesh tailored to the image content, aiming to reduce the number of mesh vertices and improve algorithm efficiency. We correct portrait images by taking faces and bodies into account to avoid inconsistent head-to-body ratios.

3. Problem Statement

In this paper, we try to correct distorted portraits in wide-angle photos. In this section, we describe the problem and the overview of our algorithm.

(a) Perspective (b) Stereographic (c) Shih *et al.*'s [27]

Figure 1. Inconsistent head-to-body ratio with only the face considered as the salient region. (a) The image rendered with perspective projection; (b) the image rendered with stereographic projection; (c) the result generated by the method of [27].

3.1. Wide-angle portrait correction problem

Projecting the 3D scene into a 2D plane using any one global projection inevitably results in distortion, such as unnatural portrait stretch or line bending. In this paper, we mainly discuss the distortions produced by perspective and stereographic projections. The perspective projection is based on the similar triangle principle. Therefore, any straight lines in the real world will be projected into straight lines in photos rendered by perspective projection. However, when the FOV is too wide, a serious stretch in the perspective image would occur in the periphery. As shown in Figure 1(a), the people in the yellow box have unnatural distortion. Different from perspective projection, stereographic projection first projects the objects into a virtual sphere, then projects the virtual point in the sphere into the imaging plane. The image rendered by stereographic projection cannot guarantee that the straight line would not be bent, but it can preserve natural portraits, as shown in red box of Figure 1(b).

Photos with stretching portraits or bending lines do not match the viewing experience of our own eyes. Therefore, we try to correct the photos rendered by one global projection. We can easily obtain wide-angle perspective photos from mobile phones with wide-FOV cameras. Thus, we focus on solving the problem of portrait distortion in the periphery of wide-angle perspective photos.

The inputs of our problem are photos with portraits taken by phones with wide-angle cameras and the corresponding 35 mm equivalent focal length f_{35}, where f_{35} is the parameter to measure the FOV of the camera. The input photos are rendered with perspective projection, where the person located in the periphery has serious stretching. The output is an image with the same resolution as the input, where the portrait distortion in the periphery of the image is corrected, and the background maintains the effect of perspective projection.

3.2. Algorithm overview

Taking the portrait-preserving of perspective projection and the line-preserving of stereographic projection, we apply stereographic projection into the human area and perspective projection into the background area to correct the wide-angle photos. Previous works focused on face correction and ignored the coordination between the bodies and the faces [27, 28], leading to inconsistent head-to-body ratios (Figure 1(c)). To overcome this problem, we select the whole portrait area for deformation. To ensure the continuity of the deformation, previous works used dense uniform quad mesh to drive the correction. The mesh vertex coordinates are regarded as variables in the least squares optimization, which measures the distortion. Therefore, too many mesh vertices would reduce the optimization efficiency. Wide-angle portrait photos are special images with clear subjects, which is why the adaptive mesh is conducive to improving efficiency by reducing vertices in the non-salient area.

We propose a novel method that corrects wide-angle perspective photos with portraits automatically. Figure 2 gives an overview of the algorithmic pipeline. Given a wide-angle perspective image (Figure 2(a)). We first generate the density map based on the portrait mask (Figure 2(b)), and then generate the adaptive triangle mesh (Figure 2(c)) based on the density map. Afterward, we perform content-aware optimization that contains the portrait, background, and boundary energy terms, obtaining the optimized mesh without distortion (Figure 2(d)). Finally, we obtain the result (Figure 2(e)) by warping Figure 2(a) using the optimized mesh. The details of the process will be described in Sections 4 and 5.

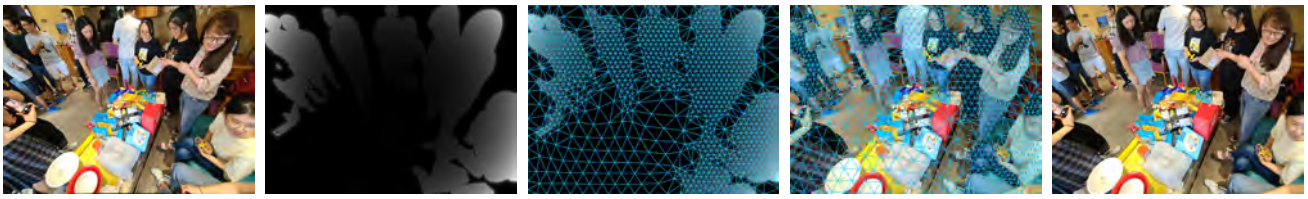
4. Adaptive Mesh Generation

Uniform quad meshes are simple but more inefficient than adaptive mesh in driving the image deformation because of their uniform dense vertex distribution. In contrast, the triangle mesh has a great potential to adapt to the image content using few vertices while guaranteeing dense vertex distribution in the salient area due to its adaptability. In this section, we generate the density map based on the portrait mask to guide the adaptive triangle mesh generation.

4.1. Density map generation

Our purpose is to correct the distortion of the portraits, which means the portrait areas should be salient. In our algorithm, we should segment the whole human area apart from the background. Instance-level human parsing [12] is the body semantic part segmentation that satisfies our requirement.

First, We obtain the initial segmentation of people as the original mask M_0. Second, we calculate the portrait mask M_t with the transition area to prevent the portrait segmentation from being insufficiently accurate. We set the brightness of portrait areas in the original mask to 100% and others to 0. The brightness is reduced the further away one is from the boundary of the portrait. The brightness calcula-



(a) Input ($f_{35} = 13mm$) (b) Density map (c) Adaptive triangle mesh (d) Correction mesh (e) Output

Figure 2. Correction pipeline.

tion equation of the transition area is as follows:

$$M_t(\boldsymbol{p}_i) = (1 - \frac{d_i}{\max\limits_{\boldsymbol{p}_j \in I} d_j})^{10}, \quad (1)$$

where d_i is the closest distance between arbitrary pixel $\boldsymbol{p}_i$ in the image I and portrait edges in the original mask M_0. For the convenience of calculation, we magnify the pixel value of M_t to 0-255 linearly. Then we obtain the portrait mask M_t with transition area by using Equation (1).

Third, to guide the generation of adaptive triangle mesh, we calculate the probability density for each pixel in the image to generate the density map M_d to adjust the vertex distribution. More severe distortion of the area corresponds to a more drastic change of the mesh vertices in the corresponding area. The mesh in distortion area should be denser to ensure the smooth transition of deformation. Because the distortion strength is proportional to the radial distance and we concentrate on the portrait distortion, we set the density map M_d as follows:

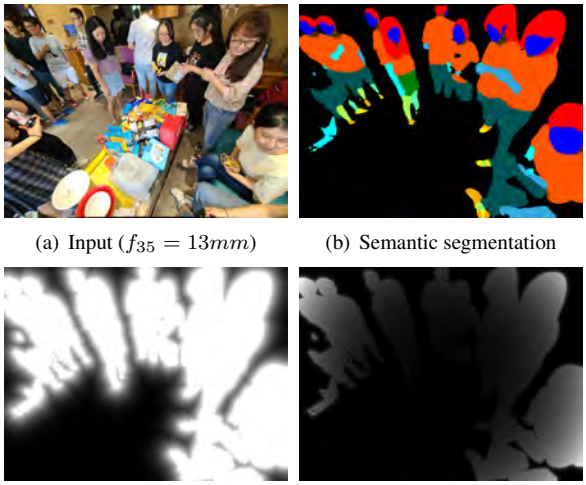
$$M_d(\boldsymbol{p}_i) = e^{\frac{\ln(10) r_i}{R_m}} e^{\ln(D_m) \frac{M_t(\boldsymbol{p}_i)}{255}}, \quad (2)$$

where r_i is the radial distance between $\boldsymbol{p}_i$ and the image center, R_m is the maximum radial distance, and D_m is the maximum probability density. In this paper we set $D_m = 1000$. Then we obtain the density map M_d by using Equation (2). The process is shown in Figure 3.

4.2. Mesh generation

We discretize the image into the adaptive triangle mesh based on the density map M_d generated in Section 4.1, such that it has denser vertex distribution in the portrait areas than background areas.

We sample the portrait area and the background area of the image with different sampling distances to obtain the initial adaptive Delaunay triangle mesh, the sampling process is shown in Figure 4. First, we sample the sparse points with background sampling distance (Figure 4(b)). Then we sample the dense points with portrait sampling distance (Figure 4(c)). Finally, we take the sparse sampling points in the background areas and the dense sampling points in the portrait areas to obtain the initial sampling points (Figure 4(d)). Assume w_{max} and w_{min} are the



(a) Input ($f_{35} = 13mm$) (b) Semantic segmentation

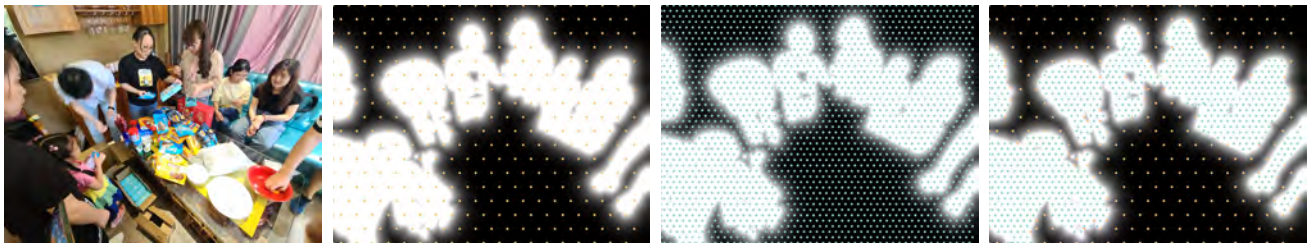
(c) Portrait mask (d) Density map

Figure 3. Portrait mask generation. Given an input i.e. perspective image in (a), we first obtain the original mask in (b) by [12]. Then we compute the transition area between the background and original segmentation to generate the portrait mask in (c). Finally, we calculate the probability density for each pixel in the image, and we visualize the density map in (d).

maximum and minimum values of the width and height of input image. We experimentally choose $\frac{w_{max}}{50}$ and $\frac{w_{max}}{20}$ to be the sampling distance for portrait and background regions. Then, we can compute that the number of mesh vertices N is in the range $(4, 25) \times \frac{200 w_{min}}{\sqrt{3} w_{max}}$.

Although simple point sampling can obtain a good adaptive triangle mesh in most cases, it will sometimes produce some sharp triangles in the transition area between the portrait and background area sometimes. We experimentally found that when the minimum angle of the sharp triangle is less than $\frac{\pi}{12}$ and the vertices cannot fit the portrait edges well (as shown in Figure 5(b)), the correction result would still be obviously distorted (as shown in the red box of Figure 5(d)).

To improve the mesh quality, by removing the sharp triangles in the transition, we use centroidal Voronoi tessellation (CVT) [23] to further optimize the initial sampling mesh. CVT updates the vertex position by using the barycentric coordinates based on the density map M_d



(a) Input ($f_{35} = 13mm$) (b) Sparse sampling (c) Dense sampling (d) Initial sampling

Figure 4. Sampling points. Inside the image, we sample points at equal sampling distance row by row, where the height between the rows is $\frac{\sqrt{3}}{2}$ times the sampling distance. And we sample points at the image boundary with the same height as the internal points.

(a) Input ($f_{35} = 15mm$) (b) Initial mesh (c) Optimization mesh (d) Results comparison

Figure 5. Initial and optimization mesh. (a) The input was taken from a mobile phone with f_{35}=15mm; (b) the initial sampling mesh, the orange points are located in the portrait areas; the green triangles are the sharp triangles; the initial mesh can not fit the contours of portraits well; (c) the CVT mesh, the orange points fit the contours of the portraits better than (b); and (d) the red box are the results generated by mesh in (b); the yellow box are the results generated by mesh in (c). After we optimize the initial triangle mesh, the sharp triangles located in the transition area between portrait and background area are eliminated.

generated in Section 4.1. After two or three Lloyd's iterations [20], the mesh quality is improved, and then we obtain the final adaptive mesh $\mathcal{M} = \{\boldsymbol{v_i}\}$ without sharp triangles in the transition area. The new mesh would fit the portrait contours well because the density map is generated based on the portrait mask with the transition area (as shown in Figure 5(c)). The result generated by optimized mesh suggest that the correction effect is improved significantly, as shown in the yellow box of Figure 5(d).

The optimized adaptive mesh can cover the salient area and transition area with high-quality mesh without sharp triangles, preventing the inner boundary of the portrait mask from being seriously affected by perspective projection, which causes the unnatural stretch.

5. Content-aware Correction

Perspective and stereographic projection have different and conflicting properties, maintaining straight line and natural faces respectively. Therefore, we apply stereographic projection to the portrait mask regions and perspective projection to the background regions. We compute the corresponding mesh $\mathcal{M}^* = \{\boldsymbol{v^*}\}$ via solving a least squares optimization problem:

$$\mathcal{M}^* = \arg\min_{\mathcal{M}} \; w_{por}E_{por} + w_{bac}E_{bac} + w_{bou}E_{bou} \;, \quad (3)$$

where $\mathcal{M}$ and $\mathcal{M}^*$ are the meshes before and after optimization, E_{por}, E_{bac} and E_{bou} are the portrait energy, background energy and boundary energy respectively; and w_{por}, w_{bac} and w_{bou} are the corresponding weights. The energy terms are designed according to the traits of the adaptive triangle mesh, which is simple, smooth and easy to optimize. The function of the three energy terms will be discussed in Section 6.1.

5.1. Portrait energy

As the input is a perspective image, we expect mesh vertices $\boldsymbol{v}_i$ in the portrait regions to be mapped into corresponding stereographic position $\boldsymbol{s}_i$, where $\boldsymbol{s}_i$ can be computed by perspective coordinates $\boldsymbol{p}_i$ using the equation from [37]:

$$\boldsymbol{s}_i = 2f \tan\left(\frac{1}{2}\arctan\frac{\boldsymbol{p}_i}{f}\right) ,$$

where f is converted by the input parameter f_{35} to describe the physical focal length using pixels as $f = \frac{R_m f_{35}}{21.6}$. Hence, we constrain the set of mesh vertices in the portrait mask to adjust positions by using the portrait energy term:

$$E_{por} = \sum_{M_d(\boldsymbol{v_i})\neq 0} \omega_i \|\boldsymbol{s}_i - (\boldsymbol{S}\boldsymbol{v}_i + \boldsymbol{t})\|_2^2 ,$$

where $M_d(\boldsymbol{v_i}) \neq 0$ indicate that $\boldsymbol{v_i}$ is located in portrait or the transition area, ω_i is the portrait weight of vertex $\boldsymbol{v}_i$, $\boldsymbol{S}$

(a) Input $f_{35} = 17$ (b) $M_w = 1$ (c) $M_w = 5$ (d) $M_w = 10$ (e) $M_w = 15$ (f) The similarity

Figure 6. The results with different M_w. (a) The input; (b⌣e) the results of corresponding maximum radical weights; (h) the similarity curve of the results corresponding to the adjacent M_w, where $D_{n,n+1}$ is the similarity between the corrected result with $M_w = n$ and $M_w = n + 1$.

is the similarity transform matrix represented by:

$$\boldsymbol{S} = \begin{bmatrix} a & b \\ -b & a \end{bmatrix} ,$$

$\boldsymbol{t}$ is a two-dimensional translation vector, which are the variables that adjust the vertex distribution through slight translation and rotation [27]. The portrait weight ω_i is set based on the distortion strength which is related to the radical distance r_i of the vertex $\boldsymbol{v}_i$. Next, we will discuss the portrait weight setting.

According to the properties of perspective projection, a greater radial distance r_i of the vertex $\boldsymbol{v}_i$ corresponds to more serious distortion. The center of the image usually does not have inaesthetic distortion, so the weight ω_i of vertex $\boldsymbol{v}_i$ should be set to be positively related to the radial distance from the center. The density map M_d that describes the probability density of each pixel satisfies the property. For convenience, we linearly scale the value of the density, and set the vertex weight ω_i as follows:

$$\omega_i = \frac{M_w}{\max\limits_{\boldsymbol{v}_j \in I} M_d(\boldsymbol{v}_j)} M_d(\boldsymbol{v}_i) \quad if \;\; M_d(\boldsymbol{v}_i) \neq 0 \, , \qquad (4)$$

where M_w is the maximum radical weight.

We now consider how the weight M_w affects the corrected results. By setting $M_w = 1, 2, \cdots$, we obtain a sequence of corrected images I_n for $n = 1, 2, \cdots$. As shown in Figure 6(a-c), the results becomes more desirable when n increases from 1 to 5. However, M_w should not be too large, otherwise other energy terms will have no effect on the result. We measure the similarity of two consecutive images I_n and I_{n+1}:

$$D_{n,n+1} = \frac{P_n \cdot P_{n+1}}{\|P_n\| \|P_{n+1}\|} \, , \qquad (5)$$

where P_n is a one-dimensional vector with each element the average of RGB values of a pixel in I_n. Figure 6(f) plots the similarity between two consecutive images I_n and I_{n+1} as n varies from 1 to 15. We can observe that the similarity $D_{n,n+1}$ is higher than 0.99 with $n \geq 5$. This implies the correction result tends to be the same when the maximum radical weight M_w reaches a specific value; see Figure 6(c-e). Based on the above observation, we take $M_w = 5$ in all our experiments.

5.2. Background energy

To maintain the structures in the background, we tailor the background energy from two aspects: keep the perspective projection in the background, and apply the user-specified geometry constraints, *e.g.* line constraint.

$$E_{bac} = E_{edg} + E_{lin} \, ,$$

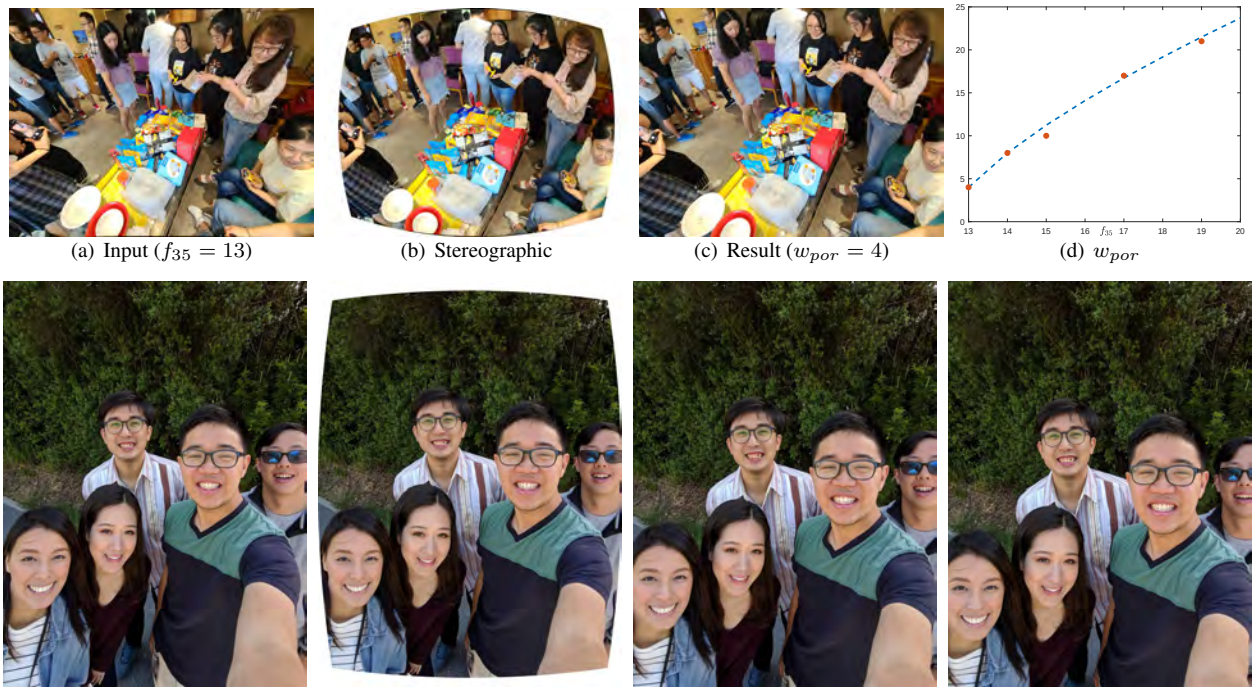
where E_{edg} is the edge term, and E_{lin} is the line-preserving term. We will discuss the two energy terms in the following.

Perspective projection preservation. We define an edge term to constraint the background of the image to maintain the perspective projection. For uniform quad mesh, the preservation of the structure benefits from the constraints of Jacobian and Hessian matrices [5, 9]. However, the Jacobian and Hessian matrices of nonuniform triangle mesh are not straightforward to estimate. As the three edges can form a unique triangle, we constraint the vector difference of the triangle edge before and after optimization to preserve the original perspective structure in the background. The term takes the length and direction message of a triangle edge into consideration, and it allows the edges of the triangle to move as parallel as possible without turning. Then the warped mesh would have a similar structure with the original mesh in the background. We can express the edge term as

$$E_{edg} = \sum_i \sum_{j \in N(i)} \|(\boldsymbol{v}_i - \boldsymbol{v}_j) - (\boldsymbol{p}_i - \boldsymbol{p}_j)\|_2^2 \, ,$$

where $\boldsymbol{v}_i$ are the mesh vertex coordinates, and $\boldsymbol{p}_i$ are corresponding perspective coordinates. Intuitively, the above energy term minimizes the vector difference of the edge before and after optimization to ensure the similarity of the original and warped triangles.

Straight line preservation. If the straight structures are close to portrait regions or covered partly by people in the



(a) Input ($f_{35} = 13$) (b) Stereographic (c) Result ($w_{por} = 4$) (d) w_{por}

(e) Input ($f_{35} = 23$) (f) Stereographic (g) Result ($w_{por} = 4$) (h) Result ($w_{por} = 17$)

Figure 7. The results with different global portrait weight w_{por} and fitted curve of w_{por}.

front, then the lines in the structures are likely to be bent or misaligned after optimization. Thus we propose a line-preserving term (6) to ensure that the points in the line are still in a straight line after optimization.

$$E_{lin} = \sum_{n} \sum_{\boldsymbol{u} \in L(n)} \|u_y - k_n u_x - b_n\|_2^2 \,, \tag{6}$$

where $L(n)$ is the vertices set in nth line; u_x and u_y are the x-axis and y-axis coordinates corresponding to $\boldsymbol{u}$, respectively; and k_n and b_n are the parameters of nth line. $\boldsymbol{u}$ can be represented linearly by the three vertices of the triangle it is in, i.e. $\boldsymbol{u} = a_1 \boldsymbol{v}_t^1 + a_2 \boldsymbol{v}_t^2 + a_3 \boldsymbol{v}_t^3$, where $\boldsymbol{v}_t^i$ $(i = 1, 2, 3)$ is the vertex in corresponding triangle t, and a_i can be computed by linear interpolation. Even though we can explicitly specify the preservation of the straight lines, in most cases, we can obtain good results without adding the line-preserving term. As long as the explicit expression of the geometric shapes can be obtained, we can incorporate the related constrains into our energy framework easily.

5.3. Boundary energy

Portrait warp may occur in the boundary of the image. Hence, fixing the boundary vertex is not an appropriate constraint. To prevent unpleasant artifacts in the boundary, boundary vertices should also participate in the optimization process. The optimized boundary vertices should be as close to the original boundary as possible. We define the boundary term as follows:

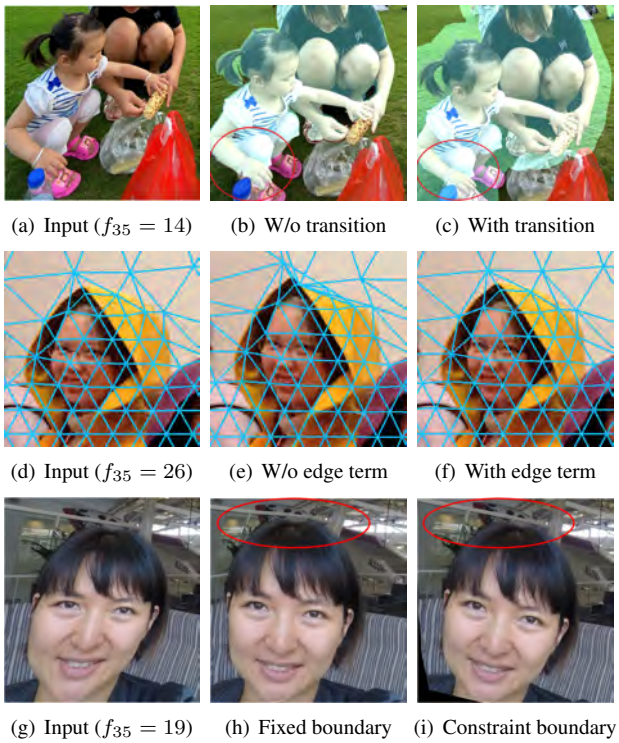
$$\begin{aligned} E_{bou} = &\sum_{i \in \mathbf{B}_l} \|v_{i,x}\|_2^2 + \sum_{i \in \mathbf{B}_r} \|v_{i,x} - W\|_2^2 \\ &+ \sum_{i \in \mathbf{B}_t} \|v_{i,y}\|_2^2 + \sum_{i \in \mathbf{B}_b} \|v_{i,y} - H\|_2^2 \,, \end{aligned}$$

where $\mathbf{B}_l$, $\mathbf{B}_r$, $\mathbf{B}_t$, $\mathbf{B}_b$ are the vertex sets of left, right, top, and bottom boundaries in the image, respectively; and W and H are the width and height of the image respectively.

5.4. Weight setting

In this section, we discuss the weight setting in Equation (3). Note that both background energy term E_{bac} and boundary energy term E_{bou} are designed to prevent distortion generation in the background and at the boundary. We also found that E_{bac} and E_{bou} are of the same order of magnitude in our experiments. For simplicity, We firstly fix the weights of energy terms to be $w_{bac} = w_{bou} = 8$, then determine the global portrait weight w_{por}.

The focal length of the photos is an essential factor affecting the distortion. Global portrait weight w_{por} that neutralizes the two projections should be adaptive according to focal length to achieve a desirable result; see Figure 7(c,g,h). We conduct experiments on images with five different focal lengths ($f_{35} = 13, 14, 15, 17, 19$) using various global portrait weight w_{por}. The optimal integer for each focal length f_{35} is denoted by $w^*_{por,f_{35}}$. We find that



(a) Input ($f_{35} = 14$) (b) W/o transition (c) With transition

(d) Input ($f_{35} = 26$) (e) W/o edge term (f) With edge term

(g) Input ($f_{35} = 19$) (h) Fixed boundary (i) Constraint boundary

Figure 8. Influence of different energy terms. (a) The input; (b&c) the results without and with transition area; (e&f) the results without and with the edge term; (h&i) the results when we fix and constraint the boundary.

the points $(f_{35}, w^*_{por,f_{35}})$ has a good fit with the function $F(x) = 2^{\ln(12.7x-157.7)}$, as shown in Figure 7(d). Thus, we set $w_{por} = F(f_{35})$ in all our experiements.

5.5. Optimization

The correction problem is formulated as a general nonlinear least squares optimization. As we described above, the energy function is a nonlinear smooth function. Therefore, we use the Levenberg–Marquardt algorithm to find the optimal. Ceres Solver is an efficient open-source C++ library for solving nonlinear least squares problems. In this paper, we use Ceres Solver [1] to solve the correction problem.

6. Implementation and Results

In this section, the ablation experiment results are first shown to describe the influence of various energy terms. Next, we show the final results of our algorithm. Then, we compare the results with the state-of-the-arts of wide-angle portrait correction [27, 28, 22]. In addition, we conduct the user study to verify user preferences on our results over results generated by other methods. Finally, we discuss the running time of our algorithm.

6.1. Ablation experiments

In this part, we will discuss the influence of the energy terms from the portrait, background, and boundary term.

Portrait energy. Portrait energy is set to eliminate the perspective distortion of the portrait areas. As shown in Figure 8, we use the masks without and with transition to demonstrate the function of the portrait terms. When we use the original mask generated from the segmentation algorithm, the little girl's right arm still has obvious perspective distortion (Figure 8(b)). After a certain transition area is added, the whole body of the little girl becomes more harmonized (Figure 8(c)). The results also present that the appropriate mask with transition can improve the effect of correction.

Background energy. We have discussed two energy terms in the background: the edge term and line-preserving term individually.

The edge term keeps the triangle mesh from flipping over and controls the similarity with the original triangle. As shown in Figure 8, the optimized triangle mesh in the hat is flipped after the optimization without the edge terms (Figure 8(e)). After adding the edge term, the foldover is avoided. As shown in Figure 8(f), we maintain a structure similar to the original mesh in Figure 8(d). Therefore, the edge term is necessary to preserve the perspective projection in the background regions.

The line-preservation term is a supplementary part of background preservation. The straight line close to or covered partly by the portrait areas may be significantly curved or misaligned after the optimization without the line-preservation term. If these types of straight lines are not processed separately, even though we have constrained the triangle edge term, the background area will still produce unpleasant deformation. As shown in Figure 9, the straight lines covered by the human body area are misaligned, and the straight lines on the ceiling above the heads are also obviously curved. After adding the line-preservation constraints, the result is improved significantly.

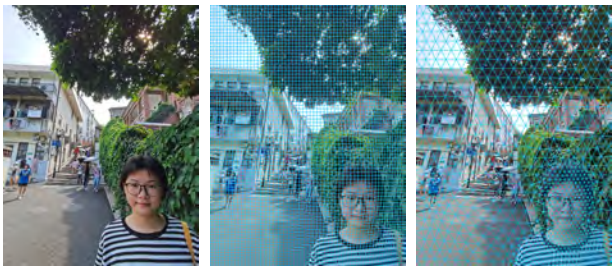
Boundary energy. Since the distortion of portraits in wide-angle photos usually occurs at the boundary of the image, if the boundary is directly fixed, the boundary does not easily produce artifacts due to the triangle edge constraint. However, the change of portrait would produce other unpleasant deformation in the transition area. For example, the edge close to the head is bent, shown as Figure 8(h). If we constrain the boundary point as close as possible to the boundaries, then the energy can be adjusted automatically, and this problem can be avoided effectively, as shown in Figure 8(i).

6.2. Corrected image results

We use linear interpolation to render the optimized images. As the boundary is free to move, the optimized image

(a) Input ($f_{35} = 19mm$) [27] (b) W/o line term (c) With line term

Figure 9. Influence of straight line term. (a) The input, and the purple lines are the straight lines specified by the user; (b) the result without constrained straight lines specified by the user; (c) the result with constrained straight lines specified by the user.



(a) Input ($f_{35} = 17$) (b) Quad mesh (c) Triangle mesh

Figure 10. Results under different mesh. (a) The input; (b&c) the result of a quad mesh with 4800 vertices and triangle mesh with 1113 vertices respectively.

would have some blank parts. we crop these areas and obtain final results. To experiment with the feasibility of the proposed method, we processed many pictures with wide FOVs, including single- and multi-person photos, and outdoor and indoor shots to ensure diverse scenes. The background of some photos contains a large amount of linear structure, as shown in Figure 11.

The experimental results show that our method is robust for a variety of scenes. For a natural background without a large number of buildings, our naked eyes cannot detect obvious background distortions. For photos with many number of straight-line structures in the background, correcting the stretching of people will cause the straight lines near the human body to be misaligned. We constrain this kind of special straight lines to guarantee a better result. Thus, our algorithm is also suitable for pictures where the structure of a building needs to be maintained. However, in most cases, we do not have to constrain all the straight lines in the picture and need to deal with only a small part of the straight lines that are bent.

To validate the robustness of our correction method, we also conduct experiments on photos taken from the same scene, as shown in Figure 12. In addition, we test our method on the photos artificially cropped according to various fields of view; the experimental results are shown in Figure 13. The results show that our method has strong robustness for images with different focal lengths.

6.3. Comparisons with the state-of-the-arts

Figures 14 and 15 show the comparisons against the state-of-the-arts of wide-angle portrait correction [37, 27, 28, 22].

Zorin *et al*. proposed a parametric method to correct the wide-angle photos [37]. As shown in Figure 14(b), the portrait has been corrected, but since the method does not process the portrait and background separately, the table and the wall in the right area are distorted. Compared with Zorin *et al*.'s method, we apply local deformation in different regions so that our result maintains the straight structures in the background (Figure 14(d)).

Shih *et al*.'s method uses a dense quad grid to drive the image deformation to guarantee the continuity of image deformation [27]. In our algorithm, the adaptive triangle mesh needs only a few vertices to achieve the same deformation effect as the denser uniform quad mesh, as shown in Figure 10. Other comparisons with Shih's work are shown in Figures 14 and 15. The results show that our method has a similar effect even though we use fewer vertices and that our method improves the optimization efficiency significantly. The optimization time will be discussed in Section 6.4.

Tan *et al*. proposed a deep learning approach based on Shih's work [28]. Although Tan's method used edge detection in the network to preserve the lines near the face boundaries, Figure 15(c) shows that the lines in the background are still misaligned when they are covered by the face areas. Tan *et al*. did not take the bodies into account; the head-to-body ratios are inconsistent as shown in Figures 14(g) and 15(h), while our result balances the head-to-body ratio as shown in Figures 14(h) and 15(j).

In addition, Nguyen *et al*. established a polynomial distortion model [22], but the straight lines between faces are

Figure 11. More results. Each group shows an input image (left), followed by our result (right).

(a) Input $f_{35} = 24$ (b) Result of (a)

(c) Input $f_{35} = 15$ (d) Result of (c)

Figure 12. The results of the photos taken from the same scene with different focal lengths. (a&c) The input images; (b&d) the results of the corresponding inputs.

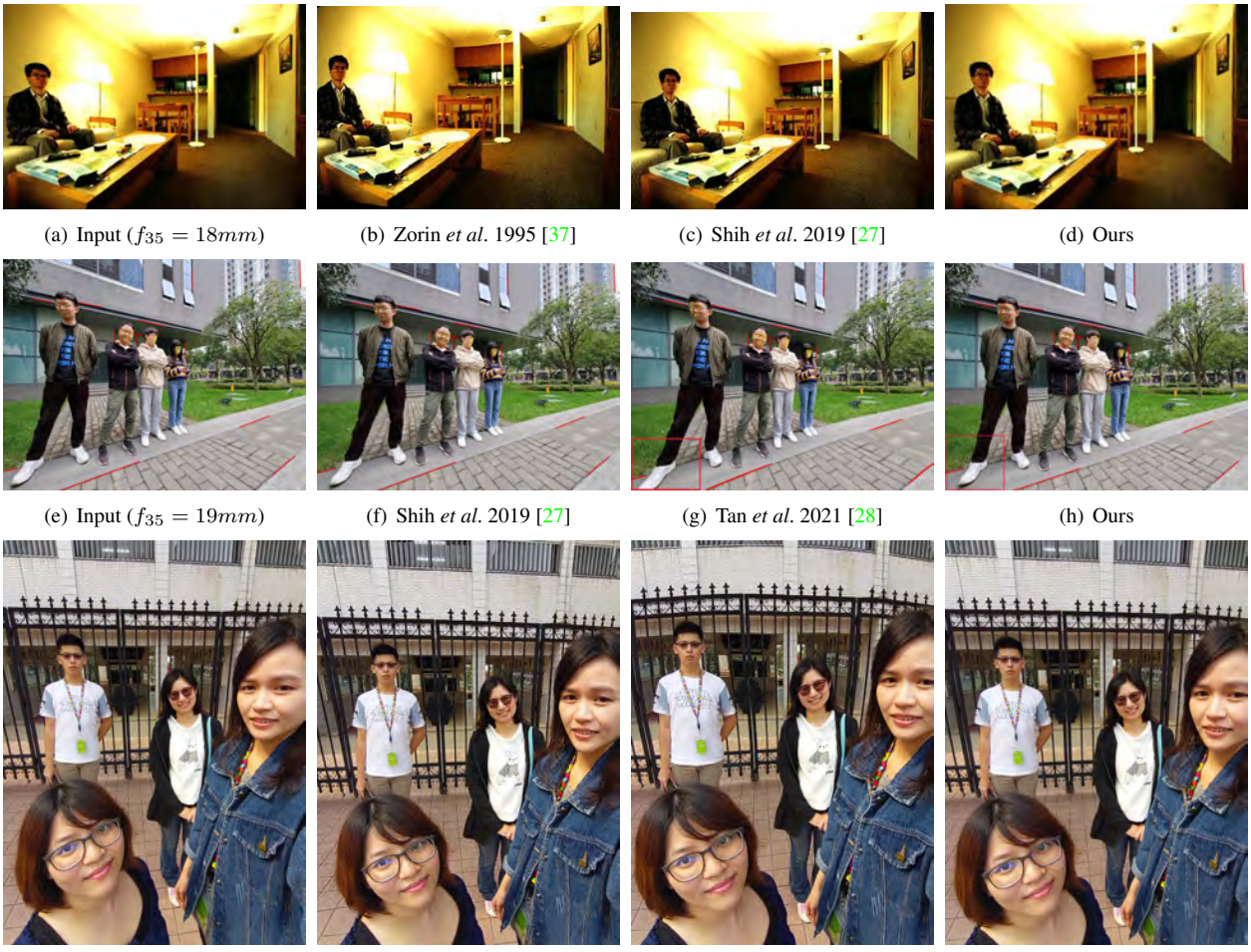
bent, as shown in Figure 14(k). Compared with Nguyen *et al.*'s model, our method has significant advantages in maintaining the background structure, as shown in Figure 14(l).

To ensure a fair comparison with other results, we first compare our results without any straight line constraints with those of the above mentioned works. However, as we consider the entire human body not just the faces, the straight edges around the salient areas more easily produce unexpected bending or misalignment after warping compared with other works (*e.g.* as shown at the bottom right corner in red box of Figure 15(i)). Straight-line perception is subjective. The straight lines detected by the existing line detection methods are messy and segmented. If the detected segmented lines are used directly in optimization, then it would lead to misalignment of the straight line after optimization and reduce the optimization efficiency. Therefore, we specify the straight lines manually to constrain (Figure 15).

6.4. Optimization time

We compare the running time of Shih's [27] and our method on a 3.20GHz Intel(R) Core(TM) i5-4460 CPU with 12GB RAM. Note that both Shih et al.'s method and our method require a body semantic part segmentation for specifying the weight in correction energies. Besides Instance-level human parsing [12], other body semantic part segmentation methods may also be applicable for our applications. Shih et al.'s approach uses a dense uniform grid to guarantee the continuity of image deformation [27], which is independent of the body semantic part segmentation result. Differ-

Figure 13. Corrections of artificially made photos. The input images are in the top row. The image on the left is the original photo ($f_{35} = 15mm$). Other images are cropped from the original image such that f_{35} becomes 16mm, 17mm, 18mm, 19mm, 20mm, respectively. The results are in the bottom row.



(a) Input ($f_{35} = 18mm$) (b) Zorin *et al*. 1995 [37] (c) Shih *et al*. 2019 [27] (d) Ours

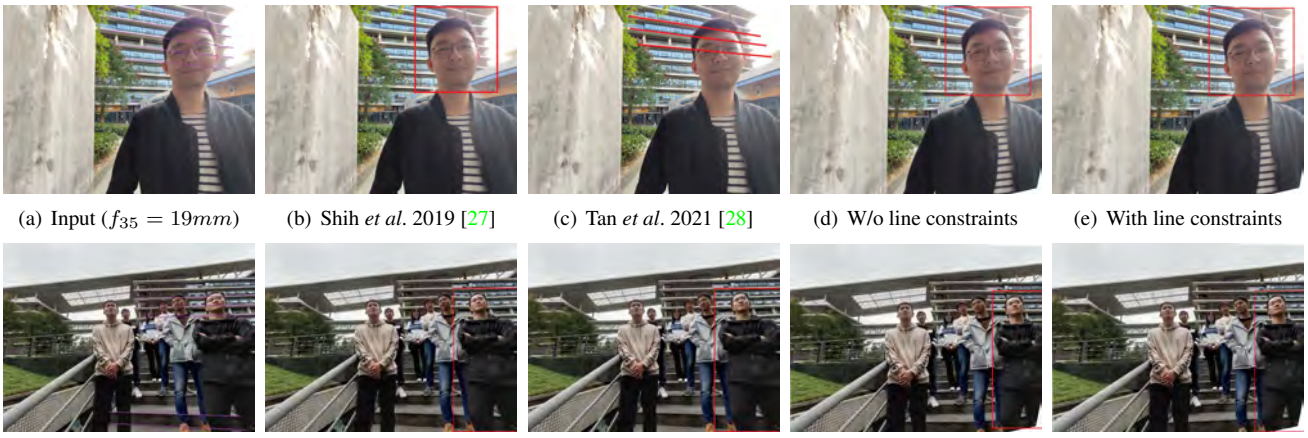
(e) Input ($f_{35} = 19mm$) (f) Shih *et al*. 2019 [27] (g) Tan *et al*. 2021 [28] (h) Ours

(i) input ($f_{35} = 19mm$) (j) Shih *et al*. 2019 [27] (k) Nguyen *et al*. 2021 [22] (l) Ours

Figure 14. Comparisons with other works.

ently, our algorithm requires a triangle mesh adaptive to the density map correlated to the body semantic part segmentation result; see Section 4. Here, we focus on comparing the computation time for mesh generation and correction optimization during the correction process, assuming an appropriate body semantic part segmentation is given.

The three sets of data in Table 1 show the number of vertices used and the running time for corresponding mesh



(a) Input ($f_{35} = 19mm$) (b) Shih *et al*. 2019 [27] (c) Tan *et al*. 2021 [28] (d) W/o line constraints (e) With line constraints

(f) Input ($f_{35} = 19mm$) (g) Shih *et al*. 2019 [27] (h) Tan *et al*. 2021 [28] (i) W/o line constraints (j) With line constraints

Figure 15. Comparisons with and without line constraints. (a&f) The input image and user-specified line constraints; (b&g) the results corrected by Shih's work [27]; (c&h) the results corrected by Tan's work [28]; (d&i) our results without line constraints; (e&j) our results with line constraints.

Table 1. Running time.

Figure	Method	N_v	T_m (s)	T_o (s)	T_t (s)
2(a)	[27]	4800	0	0.460	0.460
	Our	1176	0.030	0.045	0.075
5(a)	[27]	4800	0	0.593	0.593
	Our	1567	0.035	0.075	0.110
10(a)	[27]	4800	0	0.336	0.336
	Our	1113	0.025	0.047	0.072

Note: N_v is the number of the mesh vertices; T_m is the running time for generating mesh;T_o is the running time for correction optimization; and $T_t = T_m + T_o$.

Table 2. User study in comparisons to other methods.

	Expe. 1	Expe. 2	Expe. 3
Input	9.00%	15.00%	6.19%
[27]	N/A	16.00%	16.67%
[28]	34.00%	N/A	22.38%
Our	57.00%	69.00%	54.76%

Note: Expe. 1∽3 represent the three participant-based experiments, respectively; each col shows the percentages of votes in each experiment.

generation and correction optimization. We can observe that adaptive triangle mesh significantly reduces the number of vertices. As shown in Figure 10, the adaptive mesh uses fewer vertices and achieves similar natural faces and good background as one generated by denser quad mesh. Although extra time is used for mesh generation in our method, the overall computational time is still considerably reduced.

We also compare the optimization time of the energy function of Shih's and ours under the same number of mesh vertices. Our energy function reaches convergence faster than Shih's. For example, the optimization time of adaptive triangle mesh with 1,200 vertices is only 0.053s, while that of quad mesh with the same number of vertices is 0.083s for Figure 10(a).

In summary, our method improves computational efficiency by reducing the number of vertices and simplifying the energy function.

6.5. User Study

In this section, we further evaluate whether our results meet human expectations by conducting a user study. Here, we compared our method to mesh-based method [27] and learning-based method [28].

We presented 16 groups of images in a random order to 35 participants, including three participant-based experiments: (1) 5 groups compared with mesh-based method [27]; (2) 5 groups with deep-learning-based method [27]; and (3) the remaining 6 groups with both methods, respectively. Each group includes the randomly sampled input image with portrait distortion followed by the results generated by different methods in random order. The participants were asked to select their favorite one. Table 2 shows the distribution of votes for different methods. The voting results demonstrate that our method outperformed others; the number of votes for our results exceeds 50% in all three experiments.

7. Conclusion and Future work

We propose a content-aware portrait correction method based on adaptive triangle mesh. We build least squares optimization by minimizing the energy terms that measure wide-angle portrait distortion. Experimental results indicate that our method is robust to various kinds of pictures and that the adaptive strategy reduces the vertices significantly. Comparisons with other works demonstrate that our algorithm is efficient for optimization and the results have better consistency of heads and bodies.

One extension of our work is to realize the automatic detection of geometric shapes such as whole lines covered by human regions instead of segmented lines. A deep learning approach based on the dataset generated by our method is another potential area for future work. Inspired by [28], a worthwhile task is to add line detection to the deep learning network to achieve fully automated wide-angle correction.

References

[1] S. Agarwal, K. Mierle, and Others. Ceres solver. http://ceres-solver.org. 8

[2] O. Bogdan, V. Eckstein, F. Rameau, and J.-C. Bazin. Deepcalib: a deep learning approach for automatic intrinsic calibration of wide field-of-view cameras. In *Proceedings of the 15th ACM SIGGRAPH European Conference on Visual Media Production*, pages 1–10, 2018. 2

[3] C. Brauer-Burchardt and K. Voss. A new algorithm to correct fish-eye-and strong wide-angle-lens-distortion from single images. In *Proceedings 2001 International Conference on Image Processing (Cat. No. 01CH37205)*, volume 1, pages 225–228. IEEE, 2001. 2

[4] F. Bukhari and M. N. Dailey. Automatic radial distortion estimation from a single image. *Journal of mathematical imaging and vision*, 45(1):31–45, 2013. 2

[5] R. Carroll, A. Agarwala, and M. Agrawala. Image warps for artistic perspective manipulation. *ACM Transactions on Graphics*, 29(4):p.127.1–127.9, 2010. 2, 6

[6] R. Carroll, M. Agrawala, and A. Agarwala. Optimizing content-preserving projections for wide-angle images. *ACM Trans. Graph.*, 28(3):43, 2009. 1, 2

[7] C. H. Chang and Y. Y. Chuang. A line-structure-preserving approach to image resizing. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012. 2

[8] C. H. Chang, C. K. Liang, and Y. Y. Chuang. Content-aware display adaptation and interactive editing for stereoscopic images. *IEEE Press*, 2011. 2

[9] S. P. Du, S. M. Hu, and R. R. Martin. Changing perspective in stereoscopic images. *IEEE transactions on visualization and computer graphics*, 19(8):1288–1297, 2013. 2, 6

[10] L. Feng, M. Gleicher, H. Jin, and A. Agarwala. Content-preserving warps for 3d video stabilization. *ACM Transactions on Graphics (TOG)*, 28(3):223–231, 2009. 2

[11] R. Gal, O. Sorkine, and D. Cohen-Or. Feature-Aware Texturing. In T. Akenine-Moeller and W. Heidrich, editors, *Symposium on Rendering*. The Eurographics Association, 2006. 2

[12] K. Gong, X. Liang, Y. Li, Y. Chen, M. Yang, and L. Lin. Instance-level human parsing via part grouping network. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 770–785, 2018. 3, 4, 10

[13] K. He, H. Chang, and J. Sun. Rectangling panoramic images via warping. *ACM Transactions on Graphics (TOG)*, 32(4):1–10, 2013. 2

[14] J. Heikkila. Geometric camera calibration using circular control points. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1066–1077, 2000. 2

[15] Y. Jin, L. Liu, and Q. Wu. Nonhomogeneous scaling optimization for realtime image resizing. *The Visual Computer*, 26:769–778, 06 2010. 2

[16] J. Kannala and S. S. Brandt. A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(8):1335–1340, 2006. 2

[17] H. Lee, E. Shechtman, J. Wang, and S. Lee. Automatic upright adjustment of photographs. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 877–884. IEEE, 2012. 2

[18] T. Leyvand, D. Cohen-Or, G. Dror, and D. Lischinski. Data-driven enhancement of facial attractiveness. In *ACM SIGGRAPH 2008 papers*, pages 1–9. 2008. 2

[19] D. Li, K. He, S. Jian, and K. Zhou. A geodesic-preserving method for image warping. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015. 2

[20] S. Lloyd. Least squares quantization in PCM. *IEEE transactions on information theory*, 28(2):129–137, 1982. 5

[21] K. Nagano, H. Luo, Z. Wang, J. Seo, J. Xing, L. Hu, L. Wei, and H. Li. Deep face normalization. *ACM Transactions on Graphics (TOG)*, 38(6):1–16, 2019. 2

[22] D. H. Nguyen and T. D. Bui. An adaptive model for face distortion correction. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 2966–2971. IEEE, 2021. 1, 2, 8, 9, 11

[23] D. Qiang and F. M. Gunzburger. Centroidal Voronoi tessellations: Applications and algorithms. *Siam Review*, 41(4):637–676, 1999. 4

[24] F. Remondino and C. Fraser. Digital camera calibration methods: considerations and comparisons. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(5):266–272, 2006. 2

[25] H. Shao, Y. Wang, D. Ding, and C. Wang. Digital image aesthetic composition optimization based on perspective tilt correction. In *2020 Chinese Automation Congress (CAC)*, pages 5267–5272. IEEE, 2020. 2

[26] T. K. Sharpless, B. Postle, and D. M. German. Pannini: A new projection for renderingwide angle perspective images. In *Computational Aesthetics*, pages 9–16, 2010. 2

[27] Y. Shih, W.-S. Lai, and C.-K. Liang. Distortion-free wide-angle portraits on camera phones. *ACM Transactions on Graphics*, 38:1–12, 07 2019. 1, 2, 3, 6, 8, 9, 10, 11, 12

[28] J. Tan, S. Zhao, P. Xiong, J. Liu, H. Fan, and S. Liu. Practical wide-angle portraits correction with deep structured models.

In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3498–3506, 2021. 1, 2, 3, 8, 9, 11, 12, 13

[29] M. A. Tehrani, A. Majumder, and M. Gopi. Correcting perceived perspective distortions using object specific planar transformations. In *2016 IEEE International Conference on Computational Photography (ICCP)*, pages 1–10. IEEE, 2016. 1, 2

[30] D. Vishwanath, A. R. Girshick, and M. S. Banks. Why pictures look right when viewed from the wrong place. *Nature neuroscience*, 8(10):1401–1410, 2005. 1

[31] Y.-S. Wang, C.-L. Tai, O. Sorkine, and T.-Y. Lee. Optimized scale-and-stretch for image resizing. *ACM Transactions on Graphics (TOG)*, 27:118, 12 2008. 2

[32] J. Wei, C.-F. Li, S.-M. Hu, R. R. Martin, and C.-L. Tai. Fisheye video correction. *IEEE Transactions on Visualization and Computer Graphics*, 18(10):1771–1783, 2011. 2

[33] Y. Wu, Y.-L. Yang, Q. Xiao, and X. Jin. Coarse-to-fine: facial structure editing of portrait images via latent space classifications. *ACM Transactions on Graphics (TOG)*, 40(4):1–13, 2021. 2

[34] Q. Xiao, X. Tang, Y. Wu, L. Jin, Y.-L. Yang, and X. Jin. Deep shapely portraits. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 1800–1808, 2020. 2

[35] F.-L. Zhang, X. Wu, H.-T. Zhang, J. Wang, and S.-M. Hu. Robust background identification for dynamic video editing. *ACM Transactions on Graphics*, 35:1–12, 11 2016. 2

[36] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:1330 – 1334, 12 2000. 2

[37] D. Zorin and A. H. Barr. Correction of geometric perceptual distortions in pictures. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 257–264, 1995. 1, 2, 5, 9, 11