

Distortion-Free Wide-Angle Portraits on Camera Phones

Ji Liu Zhuoqian Yang

{jiliu, zhuoqiay}@andrew.cmu.edu



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



(b) Distortion-free photo by our method.

Figure 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.

1. Introduction

In this team project, we implemented the work from Shih et al. [2] from scratch and extended the framework to general use cases. Nowadays, wide-angle lenses have been prevailing in many mobile phone models including iPhone 12. Wide-angle lenses are able to capture a wider field of view (FOV) of the scene in the world and is thus widely used in many scenarios such as selfie and landscape. Wide-angle portrait mode has been extensively used in taking selfies because it enables the camera to include more people and faces in the photo. As a result, many modern camera phones are equipped with lenses of large FOV. For example, LG G6 has a 100° FOV front camera and a 125° FOV rear camera.

However, lenses with large FOV tend to create artifacts in the photos taken. Large distortions are introduced in the regions of large FOV. This is caused by the nature of perspective projection, which projects the surrounding world onto a flat image. These artifacts include unnatural, wider, asymmetric, and unpleasant faces in the photos, which may bring misleading impression of the included subjects. In

addition, wide-angle camera phones also suffer from fish-eye like artifacts that bend straight edges on buildings, facades, interiors, and window frames. This is a huge issue because people are taking selfies everyday and everywhere. Motivated by this issue, the authors proposed an automatic algorithm to do post-processing on-site immediately after photos are taken. Their algorithm reverses perspective distortion in wide-angle portraits, so that everyone in the photo looks natural and real, as shown in Figure 1.

Specifically, given an input image, the algorithm computes the mask around the face regions of the included subjects to balance the perspective projection and the optimized projected appearance. It also computes the stereographic projection image, a conformal mapping between a sphere and a plane, of the input perspective projection image as a pseudo-target image. Then, some energy terms are formulated to encourage face regions to locally conform to the stereographic projection for distortion restoration. The optimization process is computed on a coarse mesh over the input image to accelerate computation speed. As a result, the output image combines both the stereographic and per-

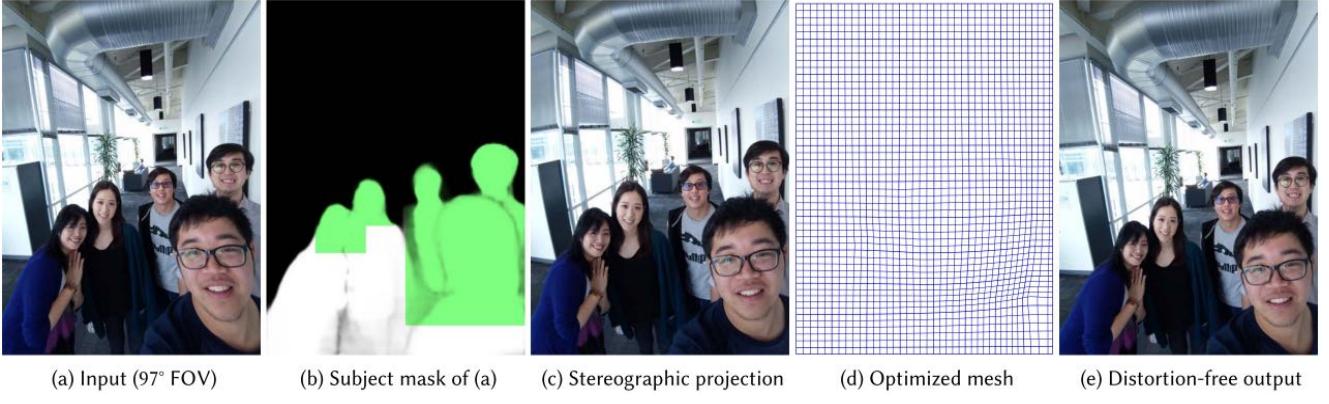


Figure 2: Overview of the approach: given a wide-angle input rendered with the perspective projection (a), it first retrieves the facial regions using a person segmentation (b), and compute the stereographic projection. In (b), we mark the region intersecting with the extended face boxes with green color. 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. 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).

spective projections on a single image. The energy function encourages smooth transitions between the two conflicting projections at face boundary. This method is designed specifically to address human face distortions because of its importance for modern mobile cameras.

In this project, we first implemented this approach from scratch while testing both on the FOV images provided by the authors and the images taken by ourselves under wide-angle portrait mode on iPhone 12. After we validated our implementation, we then extend our implementation to work on other object categories instead of being limited to human faces. We did qualitative analysis and ablation studies on the implemented algorithm to carefully study the method. In this report, we will introduce the method in detail in Section 3. In Section 4, we will do qualitative analysis and ablation study on our implementation. We will also introduce the extension to other object categories and show corresponding qualitative results.

2. Resources

- **Baseline** Our work is based on the paper [2] which provides a description of the algorithm but does not provide the implementation details such as parameters or the optimization algorithm they used.
- **Codebase** Not available.
- **Data** On [3] it provides 167 Flickr images which con-

tain large wide-angle distortion. We will use these images for evaluation and comparision. Apart from these images, we will capture some images of our own.

- **Camera** We will use the ultra wide-angle lens (13mm equivalent) on the newest iPhone to capture images for evaluation. The builtin lens correction can be disabled so that we can obtain the uncorrected images. This also enables comparison with the iPhone’s correction algorithm.

All available resources can be found on [3].

3. Method

We will describe some details of the implemented method in this section. But we will not mention every technical detail here for conciseness. Please refer to the original paper for full explanation. The whole pipeline of the implemented approach is shown in Figure 2. In the following we will briefly introduce each building block of this framework.

3.1. Stereographic Projection

Stereographic projection is a particular mapping that projects a sphere onto a plane. The projection is defined on the entire sphere, except at one point: the projection point. It is conformal, meaning that it preserves angles at which curves meet. It preserves neither distances nor the

areas of figures. Stereographic projection is desired on the face regions in our case because it can picture the 3D world onto a 2D plane with minimal conformal distortion. However, it inevitably changes the curvature of long lines, which is not desired. So we enforce the stereographic projection locally around the face regions of the included subjects to correct perspective distortion. We follow the proposed radial mapping equation in the paper to compute the stereographic projection from the input image. The equation is shown in Figure 3.

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

Figure 3: Radial mapping equation used to compute stereographic projection. 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. f is the focal length of the camera.

3.2. Subject Mask Segmentation

The focus of this method is to correct the face and hair regions for all the subjects, while leaving the background untouched. So we compute the masks of face regions of the included subjects and enforce stereographic projection locally on the face regions we want to undistort without introducing artifacts to the background. But different from the off-the-shelf subject segmentation network and the facial landmarks used in the original work, we use Detectron2 [4] to extract subjects’ masks and use Python’s Dlib to extract face bounding boxes. Then we take the intersection of the two to get our final face mask. We empirically validate our mask results by visualizing the masks overlayed with the input images.

3.3. Naive Blending

Now it is time to combine the results of the stereographic and perspective projections into a single optimal image. Our purpose is to only correct the local face distortion using stereographic projection while leaving the straight lines in the background remain being perspective projection. The most intuitive way is to use the face masks to naively blend the two projections into a single one. Specifically, we can use stereographic projection only in the face regions and use perspective projection in the background regions. However, this naive blending can introduce severe artifacts around the face, which is not acceptable. This is because of the large misalignment between the two projection types. Therefore, energy minimization approach is used to reconcile the con-

flicts between these two projections at face boundaries, as introduced in the following sections.

3.4. Local Face Undistortion

We will introduce four terms used in the optimization process, namely Face Objective Term, Regularization Term, Line-Bending Term and Mesh Boundary Term.

3.4.1 Face Objective Term

The equation of Face Objective Term is shown in Figure 4. We compute a Face Objective Term for each face and the energy of all faces will be aggregated together to form the final Face Objective Term.

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

Figure 4: Equation to compute the per-face energy $E_{s,k}$, where w_i are the face weights in the face masks, \mathbf{u}_i are vertices on the stereographic mesh, and B_k denotes the set of vertices on the k -th face. m_i are the weights sampled from a radial sigmoid function to account for the stronger perspective distortion at the image corners. The radial sigmoid function leaves the image center unaffected and free from fish-eye artifacts at the center of stereographic projection. S_k and t_k represent the per-face similarity transform, which is designed to facilitate the optimizer to find a better vertices arrangement by slightly translating, rotating, and scaling each face individually.

3.4.2 Regularization Term

The Regularization Term is used to encourage the displacement between the center vertices and their neighbor vertices to be small, as shown in Figure 5.

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

Figure 5: Regularization Term used to encourage smoothness between 4-way adjacent vertices.

3.4.3 Line-Bending Term

The Line-Bending Term is used to prevent the shearing and rotation artifacts introduced in the optimization process. This issue can happen especially on the boundary between

the face and the background, because the two regions follow different projections. This term encourages the direction of the lines to be preserved. The equation is shown in Figure 6.

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

Figure 6: Line-Bending Term designed to prevent twist, where e_{ij} is the unit vector along the direction $p_i - p_j$, 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.

3.4.4 Mesh Boundary Term

Finally, we enforce the mesh boundary constraint which preserves the pixel values on the mesh boundary to avoid the trivial null solution. We also expand the mesh boundary by a small amount to account for cases where faces are close to image boundary.

3.5. Implementation Details

Instead of using a fixed mesh dimension for all images, we use a linear scale factor, i.e. the mesh dimension is defined as $[W/40] \times [H/40]$, where H and W are the height and width of the input image. Our implementation is tested in a desktop setup with a AMD 3900 CPU and a Nvidia 1080 Ti GPU. The optimization framework we use is PyTorch. We run 200 iterations with an Adam [1] optimizer with learning rate 0.5. Inference time is close to that reported by the authors. We implemented the algorithm with a total of 531 lines of code.

4. Experiments

4.1. Ablation Study

In this section, we present an ablation study showing the effect of each part of the algorithm we implemented. The data in Fig. 7a is captured by us. The rest of the test images and corresponding FOV values are obtained from the paper's supplementary materials. We show the effect of each component by progressively adding them to the pipeline. We've selected the images such that the effect of the components of the algorithm can be visualized. In these results, the first row is the input image. The left column shows the optical flow that is used to warp the input image, the second column shows the outcome of the warping. The component added at each row are:

1. **Stereographic Correction.** Stereographic projection shows the image without perspective aberration. The images obtained after stereographic correction has undistorted face shapes, but other parts of the image are distorted.

2. **Subject Mask.** Subject masks are used to localize the correction to only subject areas such that the background is not affected. However, simply limiting the correction to these areas causes artifacts in the boundaries of these areas. This is visible in all of the cases.

3. **Regularization.** In order to obtain a smooth warped image, the regularization term is added. The term penalizes large distance between vertices and moves vertices that are neighbor to the subject area.

4. **Line Bending Term.** A simple regularization restricting inter-vertex distances causes straight lines to be curved, this is visible in Fig. 7a. To preserve straight lines, the line bending term is introduced.

5. **Mesh Boundary Extension.** In Fig. 7b, the face is located close to the border of the image and caused artifact when the boundary of the image is warped. To prevent this kind of artifact, mesh boundary extension is introduced. This term constrains the border of the image to align with the input, similar with what we do integrating an image processed in gradient domain.

6. **Similarity Transformation Constraint.** Some times the stereographic correction enlarges the size of the head, such as in Fig. 9a. This is undesirable. To preserve the size of the faces and prevent other unusual distortions, the similarity transformation constraint is introduced. This term constrain the transformation in each facial area to be a similarity transformation with scale factor $s = 1$.

4.2. Qualitative Results

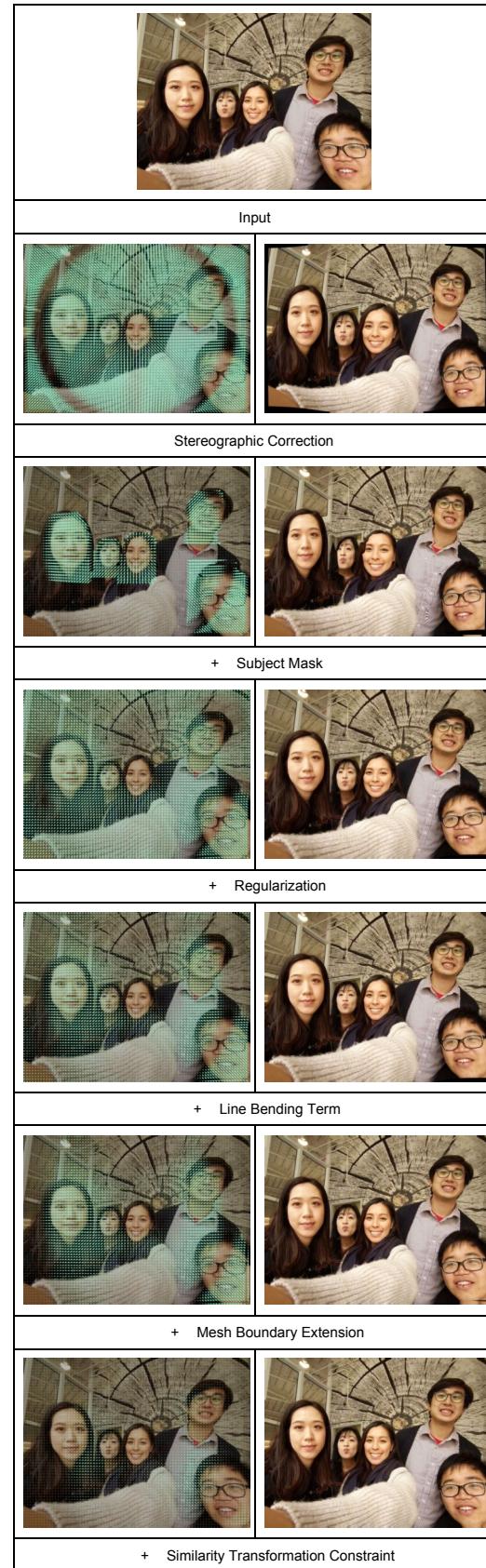
More results on Flickr images with different FOVs are shown in Fig. 10. Our implementation is able to generate reasonable correction of the human faces without distorting the background. In these results, the first column shows the input image, the second column shows the optical flow that is used to warp the input image, the third row shows the output.

4.3. Extension to general objects

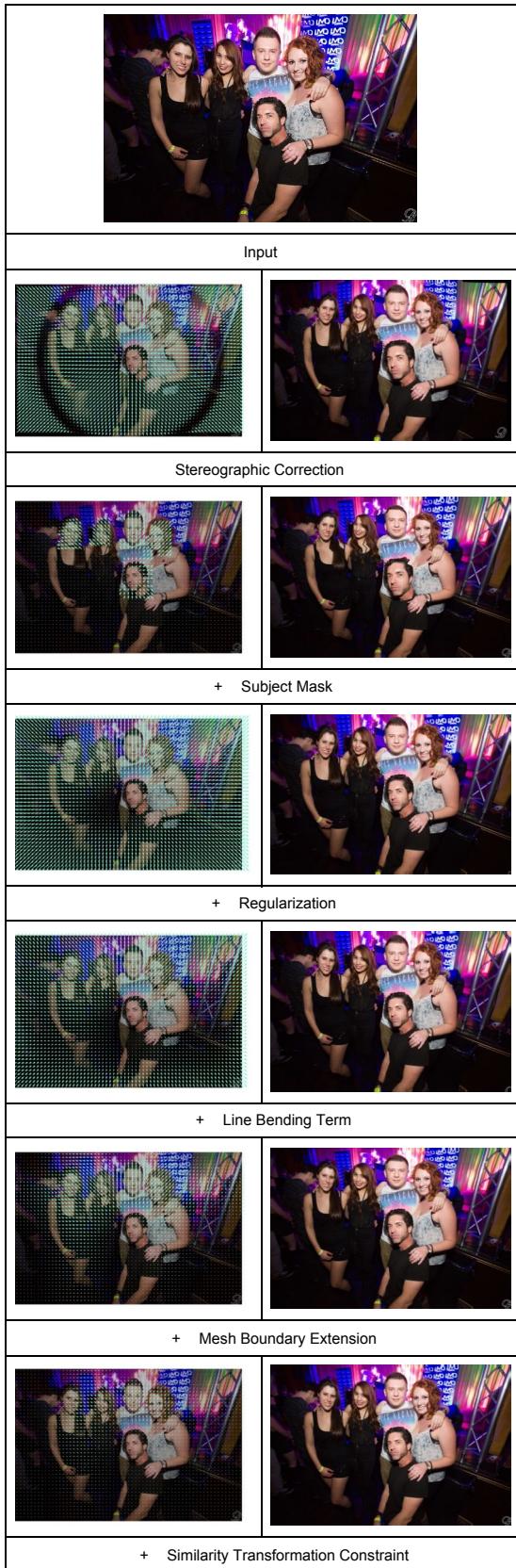
We extend the method to correct aberration in not just human faces but also other general objects such as cars, bicycles, full human bodies and animals. The extension is implemented by using a wider range of object categories when performing subject segmentation. We tweaked the parameters so that it will work better for general objects. The



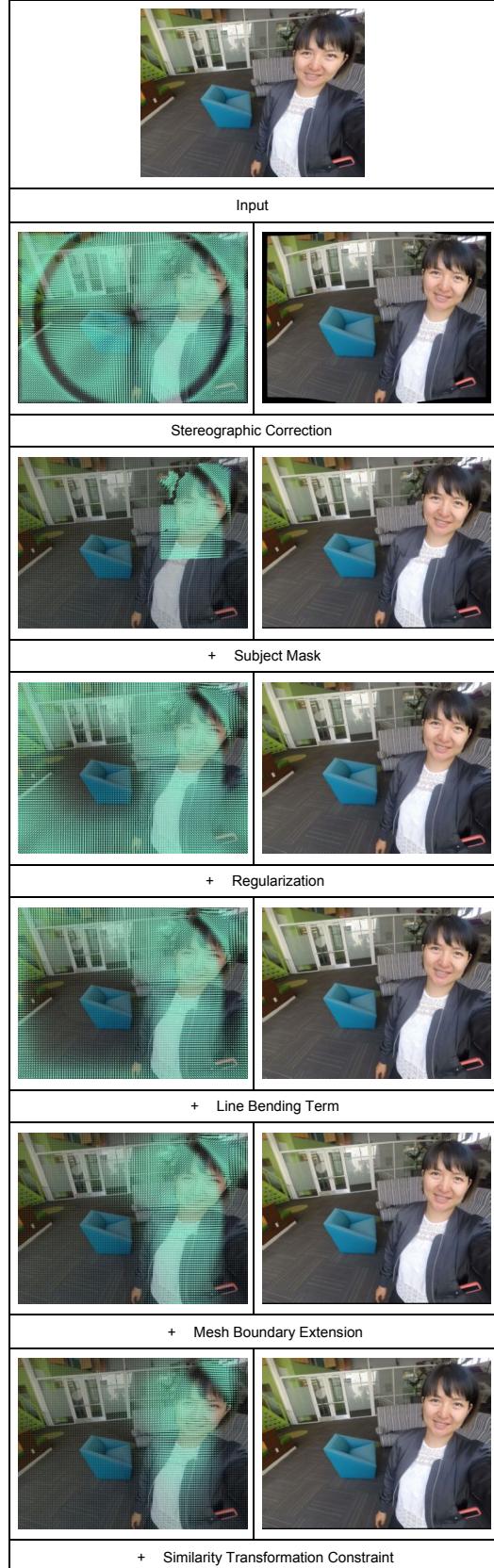
(a) Ablation Study Case #1



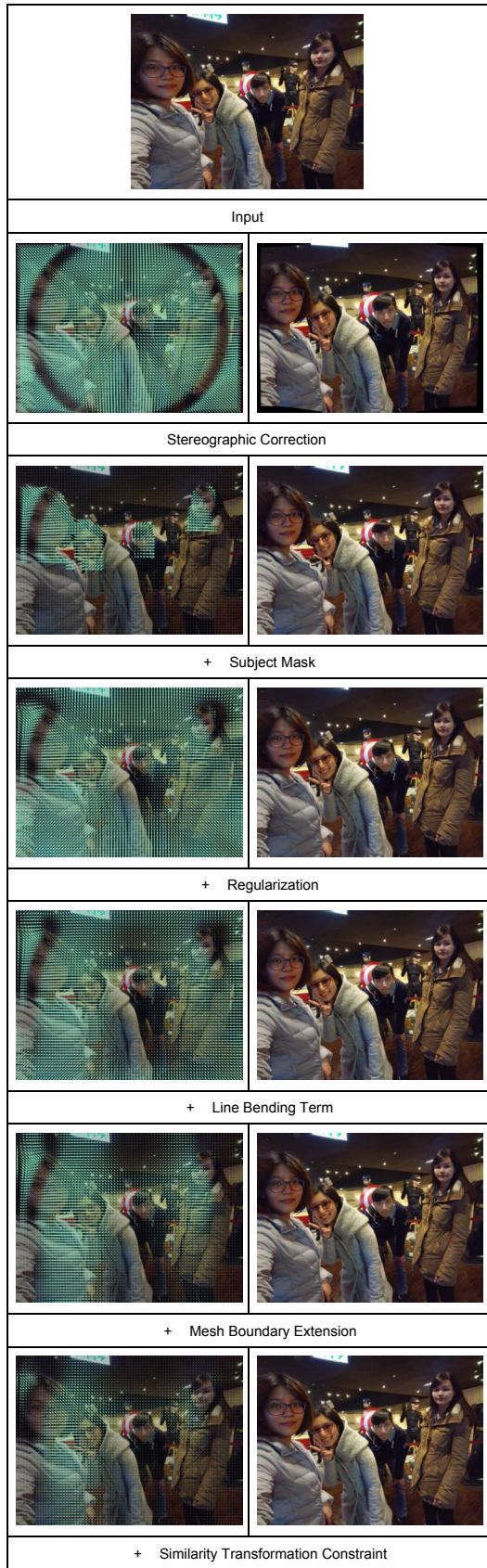
(b) Ablation Study Case #2



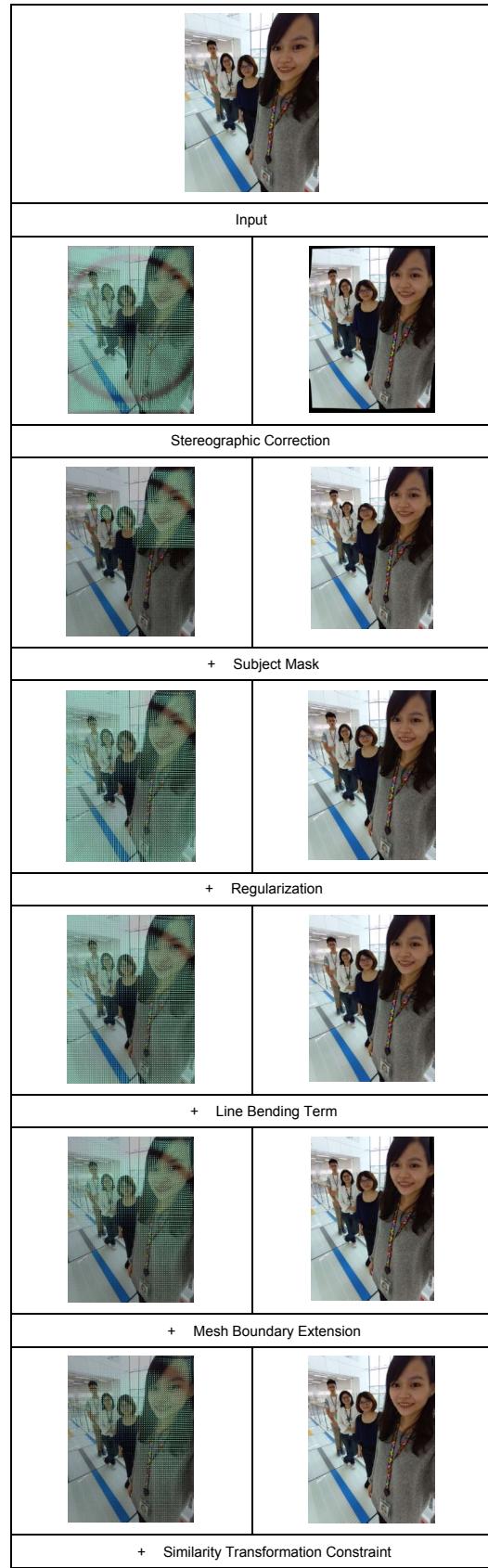
(a) Ablation Study Case #3



(b) Ablation Study Case #4



(a) Ablation Study Case #5



(b) Ablation Study Case #6

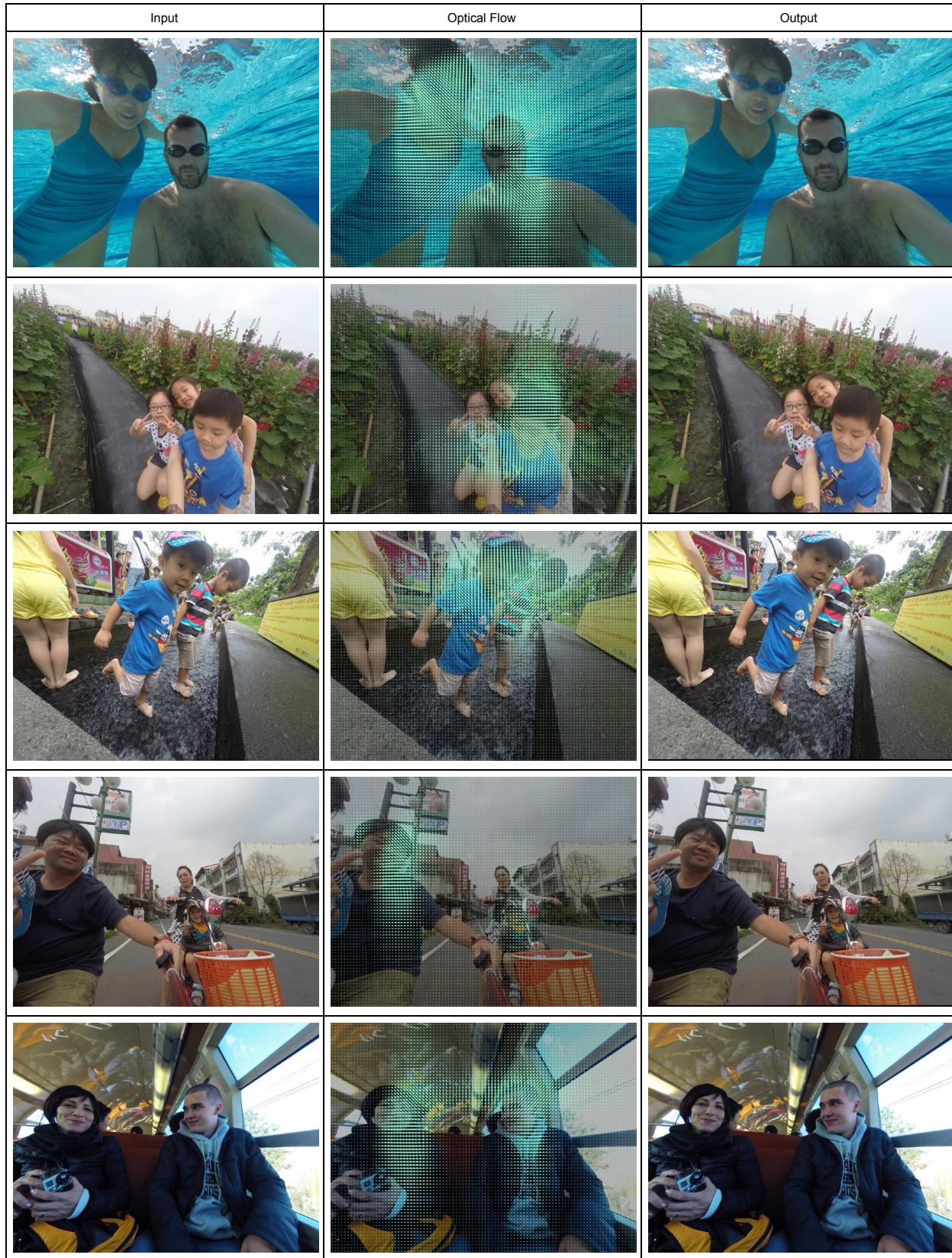


Figure 10: Qualitative Results. The first column shows the input image, the second column shows the optical flow that is used to warp the input image, the third row shows the output.

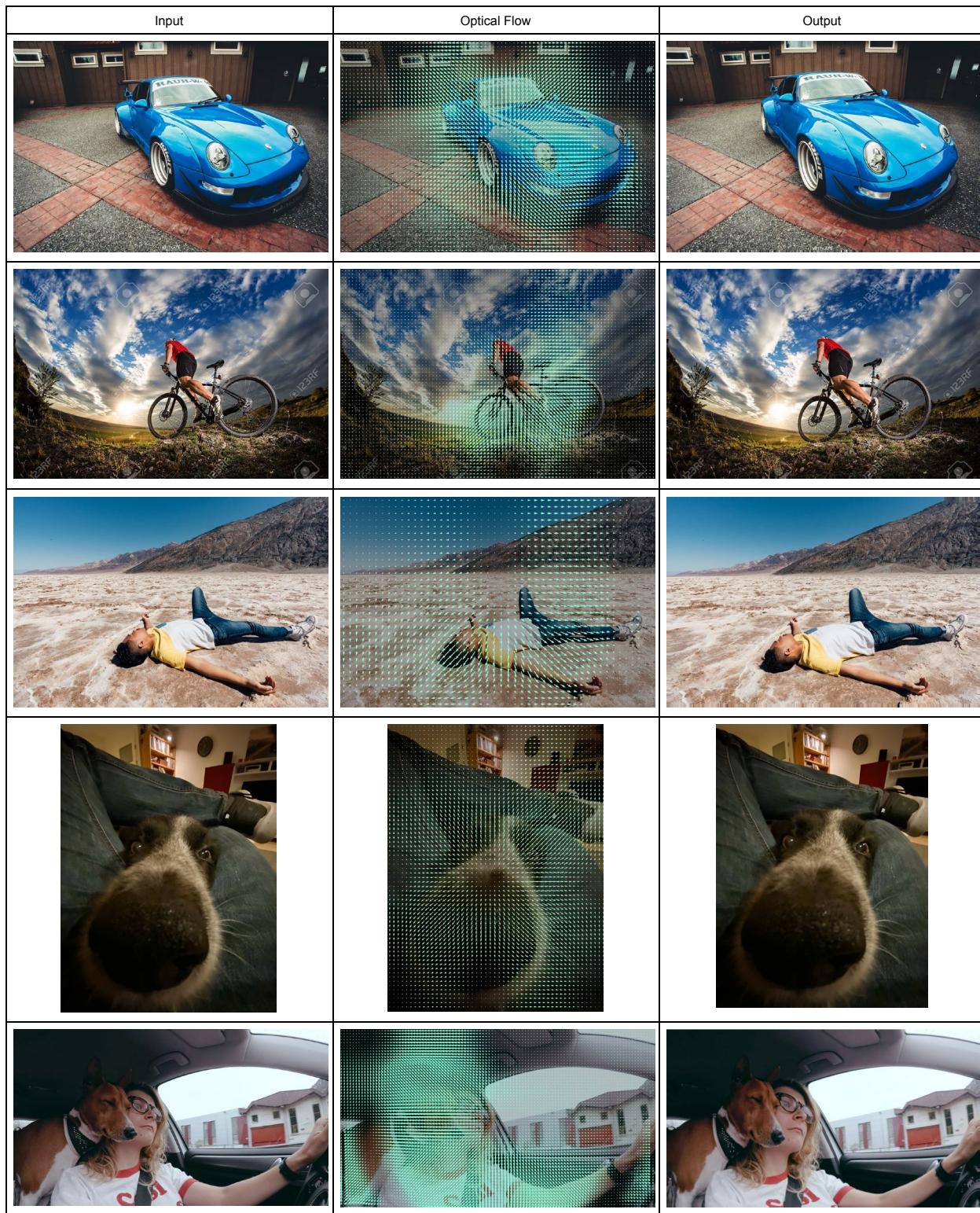


Figure 11: Extension of the method to general objects

results are shown in Fig.11. The extended method is able to correct subject areas of the image to closer to human perception. We notice that although the corrected results shows less perspective aberration, there is no significant aesthetic improvement or in other ways enhance the image. The reason could be that (i) When taking a portrait, especially a group portrait, the face of the subjects are almost facing the camera. In such a circumstance, the faces can be approximated as circular flat surfaces that are orthogonal to the camera rays, which makes orthographic correction feasible; (ii) human perception is sensitive to human faces and the distortion of it. As for other objects, the perspective distortion does not result in unpleasant feelings and is sometimes used in favor of dramatic expression.

References

- [1] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 4
- [2] YiChang Shih, Wei-Sheng Lai, and Chia-Kai Liang. Distortion-free wide-angle portraits on camera phones. *ACM Trans. Graph.*, 38(4), July 2019. 1, 2
- [3] YiChang Shih, Wei-Sheng Lai, and Chia-Kai Liang. Project page - distortion-free wide-angle portraits on camera phones. http://people.csail.mit.edu/yichangshih/wide_angle_portrait/, 2019. 2
- [4] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. Detectron2. <https://github.com/facebookresearch/detectron2>, 2019. 3