MaDCoW: Marginal Distortion Correction for Wide-Angle Photography with Arbitrary Objects

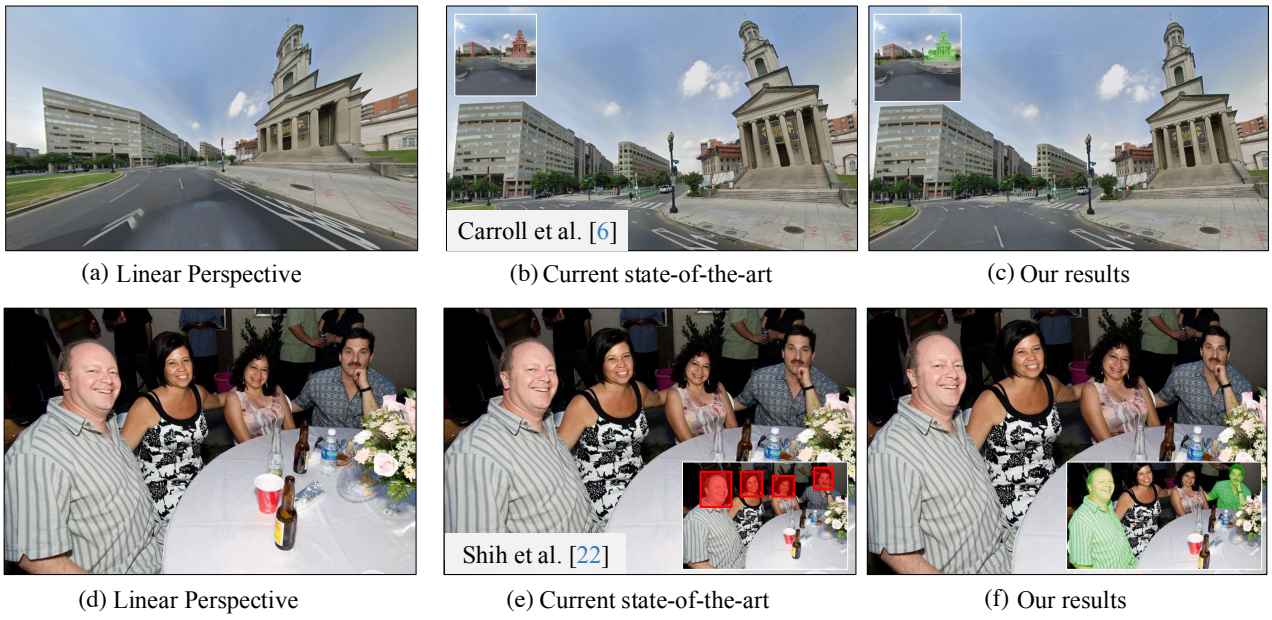
Kevin Zhang
UMD

Jia-Bin Huang
UMD

Jose Echevarria
Adobe Research

Stephen DiVerdi
Adobe Research

Aaron Hertzmann
Adobe Research



(a) Linear Perspective (b) Current state-of-the-art (c) Our results

(d) Linear Perspective (e) Current state-of-the-art (f) Our results

Figure 1. **Correcting marginal distortion in arbitrary objects.** (a) Linear perspective projection of a wide field-of-view (here, 150°) produces extreme marginal distortions, such as the buildings' distorted appearances. (b) The method of Carroll et al. [6] with line annotations (top left inset) preserves some straight lines, but the building on the right seems to lean left. (c) For MaDCoW, we supply a single region annotation for the building on the right, and MaDCoW successfully avoids distortion for the whole building. (d) A linear perspective photograph (90° field-of-view), from Shih et al. [26]. The man on the left appears widened. (e) Shih et al. [26] correct distortion on faces, but not on bodies. The leftmost person's head is corrected but not his body, making his body look even larger, relative to his head. (f) MaDCoW makes people on the margins appear more realistic.

Abstract

We introduce MaDCoW, a method for correcting marginal distortion of arbitrary objects in wide-angle photography. People often use wide-angle photography—it is the default in smartphone cameras—but very-wide-fields-of-view produce distorted object appearance in image margins. In our system, a user annotates straight lines and regions of interest. MaDCoW solves for a separate linear perspective projection for each region and then jointly solves for a distortion-minimizing projection for the whole photograph. We show that MaDCoW can produce good results in cases where previous methods yield visible distortions.

1. Introduction

To convey large spaces, people often use wide-angle photography; today's smartphones shoot wide-angle by default. Wide-angle refers to photographs where the camera's field-of-view is much wider than how people normally view pictures [10]; this mismatch in viewing condition causes shapes to appear distorted. Notably, *marginal distortion* (Figure 1(a)) occurs when objects appear in the sides or corners of a wide-angle linear perspective picture. To avoid distortion, many artists and scientists have developed alternative wide-angle projections [13]. However, alternative projections typically create other kinds of distortions, such as bending straight lines, and no single projection can eliminate all distortion [35]. As a result, state-of-the-art methods

adapt to image content, adjusting projections in the presence of specific kinds of features such as straight lines [6], faces [26], rectangular rooms [31], and fronto-parallel rows of buildings [1, 23, 24]. However, each method is specialized for very specific situations, and none can correct distortion for arbitrary objects or general image regions.

This paper introduces a technique for creating wide-angle images to correct perspective distortions for image regions containing *arbitrary objects*. Given an input photograph, the user labels one or more regions-of-interest (ROIs) for undistortion, as well as straight lines to be preserved. Our algorithm then optimizes a reprojection of the photograph to achieve these goals, while also maximizing conformality in a content-aware manner [6].

Our key idea is to employ the Direct View Condition (DVC) [13, 35]: *objects appear undistorted if and only if they appear as they would near the center of a linear perspective projection*. For each ROI, we seek a linear reprojection to removes distortion. However, the obvious choice of a linear camera pointed directly at the ROI center may produce a mismatch with the object's background. We thus optimize the projection to minimize these discrepancies.

The main steps of our approach are as follows (Figure 2). Given a photograph with user-provided line and ROI annotations, we first optimize a linear perspective projection for each individual region. We then perform a combined optimization over the entire projection to match these projections, while preserving straight lines and maximizing conformality.

The contributions of this paper are as follows:

- Distortion correction of regions containing general objects, by optimizing projections for each ROI.
- Since different users may have different goals, we allow users to author the projections by specifying which regions are most important for undistortion.
- An argument that the conformal loss [6] can be derived from the DVC (Appendix A), as can straight-line constraints. Hence, we pose projection authoring as selectively applying the DVC to different image regions.

2. Related Work

When linear perspective was invented in the Renaissance, artists quickly discovered the problem of *marginal distortion*, in which objects appear stretched in the corners of pictures (Figure 1(a)). Leonardo da Vinci recommended against applying consistent linear perspective for objects in picture margins [16]. And, no classical painter seems ever to have painted faces with marginal distortion, despite the long history of wide-angle paintings [13]. Many painters throughout history have experimented with different approaches to perspective in order to avoid the shortcomings of linear perspective [15]. Yet, our everyday photography continues to rely on wide-angle linear perspective.

Computational photography research has developed many techniques for producing wide-angle photographs while avoiding perceived distortions [13], like marginal distortions and straight lines becoming curved. Many methods involve user control to specify goals and constraints.

These projections generally have one or more goals, including: avoiding apparent distortions (e.g., preserving straight lines, avoiding stretching faces), matching the input picture; and/or user-specified constraints. However, as Zorin and Barr showed [35], no single projection can achieve all desirable goals simultaneously, and so compromises are achieved by blending separate content-specific projections. Most methods require some interactive guidance for a user to specify some of their goals. Existing methods can work well for specific content types, but effective authoring of wide-angle projections for arbitrary content remains an open problem.

Zorin and Barr [35] proposed a curvilinear projection that compromises between reducing marginal distortion and preserving straight lines. Agrawala et al. [2] proposed a computer graphics framework for multiperspective projection, in which each object is rendered with separate view directions. One class of techniques works for street scenes, containing long rows of building façades that are all roughly fronto-parallel [1, 23, 24]. Perspective warp [7] allows adjustment of a single-perspective building, and Lee et al. [18] describe a method for automatically making a single building more upright.

Our paper focuses specifically on the projection of a wide-angle view constructed from a single viewpoint. Panorama stitching methods, like that of Chugunov et al. [9], construct wide-angle views but do not preserve shape (e.g., straight lines) across views [27]. Zelnik-Manor et al. [31] proposed concatenating multiple linear perspective projections to reduce marginal distortion. Shih et al. [26] describes a method solely for correcting marginal distortion of faces using stereographic projection. Their method is fully automatic and efficient but only corrects face distortions and introduces other distortions in the process, e.g., for straight lines adjacent to faces. Zhu et al. [34] trains a transformer to correct marginal distortion of faces while preserving straight lines, and Lai et al. [17] extend face correction for video processing. Tehrani et al. [28] describe an interface where a user interactively adjusts the projection of an individual object or person, which is then composited back into the picture. Alzayer et al. [3] train a model that enables editing images in photorealistic ways, by fixing distortions due to image editing.

Most closely related to our work, Carroll et al. [6] take user-specified straight-line annotations and face detection inputs and optimize an image warp according to these sparse constraints. The optimization attempts to preserve straight lines while maximizing conformality in other re-

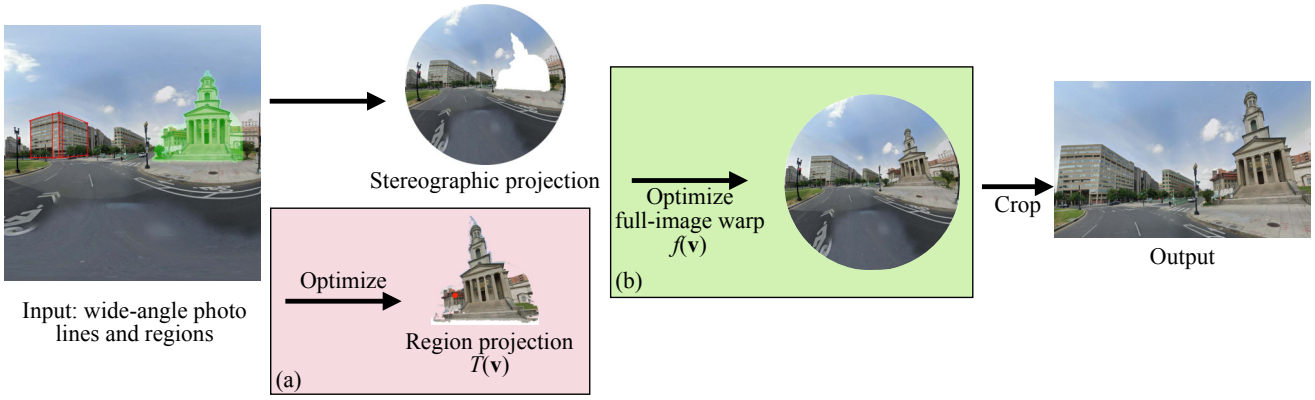


Figure 2. **Method overview**. MaDCoW takes as input a photograph, with region-of-interest (ROI) and line annotations, and outputs a wide-angle photograph with minimal distortion on the ROIs and straight lines. The algorithm has two main stages. (a) For each ROI, the method optimizes a linear projection $T(\mathbf{v})$ to best fit the shape of that region in a stereographic projection. (b) In the second stage, the method optimizes a full-image warp to match the linear projection while preserving any annotated straight lines, and maximizing conformality and smoothness. This produces an image that can be cropped into a rectangular shape if desired.

gions, especially textured regions. More recently, Zhang et al. [32] extends the work of Carroll et al. [6] by using deep neural networks to automatically detect the straight lines to be preserved and salient regions for which conformality should be maximized.While these methods can work well for preserving the appearance of lines and faces, distortions frequently remain in the spaces between annotations. We propose removing distortion for general objects by identifying separate linear perspective projections for user-annotated regions while integrating them with a whole-image warping optimization.

Nonlinear projections have been developed for other goals, including adjusting sizes of distant objects [4, 5, 21, 25], producing a rectangular picture border [8], correcting depth distortion in faces [11], making buildings appear vertical [18], artistic multiperspective images [30]. Many other methods compute linear perspective corrections to optics with radial distortion, e.g., [14, 19, 20, 22, 29].

3. Problem Statement and Parameterization

Our method takes as input a wide-angle photograph, together with a set of straight-line and region annotations, and produces an output photograph designed to satisfy those goals while minimizing distortion. No single projection works best for all scenes, and different users may have different goals, so user input is necessary—and desirable–to determine which constraints to employ.

Parameterization. The input photograph measures light on a subset of directions on a view sphere (Figure 3), where $\mathbf{v}$ is a ray direction emanating from the photograph's center-of-projection. We use 360-degree photography as inputs to capture very wide fields-of-view, but the same parameterization can be used for any type of photograph (as we do in Figure 1, bottom row).

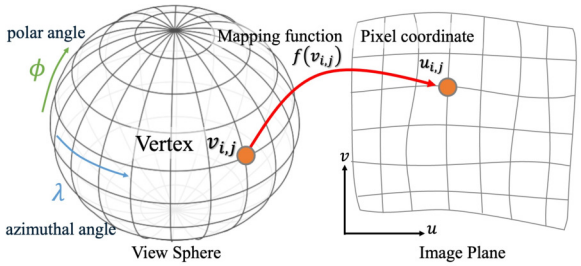


Figure 3. Viewing and warping parameterization (Figure from Carroll [6]). The view sphere parameterizes directions $\mathbf{v}$ from the center-of-projection (COP) by angles (λ, ϕ). These angles are discretized to produce a quadrilateral mesh with vertices $\mathbf{v}_{i,j}$. Each vertex maps to image space coordinates $\mathbf{p}_{i,j}$, and the full mapping $f(\lambda)$ is bilinearly interpolated within each quadrilateral.

We seek a mapping from the view sphere to the output image, defined by $f(\mathbf{v}) = \mathbf{p}$, where $\mathbf{v}$ is a direction on the view sphere, and $\mathbf{p}$ is an output image coordinate. This mapping is parameterized by a quadrilateral mesh, defined by uniformly discretizing the view sphere in latitude and longitude (Figure 3) [6]. Vertex (i, j) corresponds to ray direction $\mathbf{v}_{i,j} = (\lambda_{i,j}, \phi_{i,j})$, and gets mapped to $\mathbf{p}_{i,j}$; these $\mathbf{p}$ values are unknowns in the optimization, and f is defined by bilinear interpolation over the quadrilaterals. We aim to determine the $\mathbf{p}$ values to reproject the photograph.

Initial mapping. We focus on very wide-angle photography, where linear perspective would produce extreme distortions (Figure 1(a)). For this reason, we initialize the pro-

cess by applying a stereographic projection, to produce initial values of $\mathbf{p}$, since it does not produce marginal distortion, though it also does not preserve straight lines. Stereographic projection has desirable property of *conformality*, which makes individual triangles appear as though viewed directly [6, 15]; we argue in Appendix A that conformality can be derived from the Direct View Condition.

4. Distortion Correction for a Single Region

We begin by describing how we determine target projections for individual image regions. Later in Section 5, we describe an image warp to adjust the rest of the image accordingly.

Projecting a wide-angle scene by linear perspective produces marginal distortion (Figure 4(a)). However, if we reproject the photograph so that the *principal point* lies at the center of the target object (Figure 4(b)), then that object no longer appears distorted. The principal point is defined as the point where the camera's optical axis meets the picture (Figure 4(a)); normally, it lies in the picture's center, but moving the principal point allows us to control which objects appear distorted.

This observation motivates the use of the *Direct View Condition* [13, 35]: we can avoid distortions by using local linear projections centered on objects. One strategy to avoid distortion would be to partition the photograph into smaller regions, and reproject each to have its principal point at the center [31], but this does not produce continuity across image regions, e.g., straight lines would bend. Applying conformal mappings to individual regions [6, 26] does not preserve straight lines either. Instead, we use a linear perspective projection for the region, but we need to carefully select which linear projection.

4.1. Optimizing a single object

The obvious choice for reprojection—linear perspective with a principal point at the region's center—produces an undistorted appearance (Figure 4(b)). We also want the object's outline to fit the background. We can use a stereographic projection as a background since it avoids marginal distortion (Figure 5). However, the outline of the object may still not fit well with the background, both because both the shape of the outline and the image-space orientation in the image have changed. But the DVC says that any principal point on or near the object will appear undistorted (Figure 4(c)). Moreover, we can rotate and scale the object without affecting distortion.

Hence, given a region-of-interest mask, we optimize a new linear perspective projection for that region so that it appears undistorted while also fitting its boundary in the target photograph under stereographic projection.

Projection parameterization. We parameterize the projection by the camera's optical axis $\mathbf{v}_P$, and an affine image-space transformation. The image-space transformation includes uniform scaling and rotation in image space, parameterized by a similarity matrix $\mathbf{S} = [a, b; -b, a]$ and translation $\mathbf{t} = [t_x, t_y]^T$. Then, the projection from the view sphere to the image plane is given by:

$$T(\mathbf{v}) = \mathbf{S}\, P(\mathbf{v}) + \mathbf{t} \tag{1}$$

where $P(\mathbf{v})$ corresponds to a perspective projection that has optical axis $\mathbf{v}_P$, and the same center-of-projection as the input. We derive $P(\mathbf{v})$ in the Supplemental Material.

Optimization. We minimize the difference between the object's boundary under linear perspective, and the region boundary in the initial stereographic projection (Figure 5). We hold locations $\mathbf{p}$ for vertices outside the object fixed as their initial stereographic projection, whereas the vertices $\mathbf{p}$ within the mask are determined by the new projection. We optimize using the **conformal loss** [6] along the boundaries:

$$\begin{aligned} E_c = &\sum_{(i,j)\in V} w_{i,j}^2((v_{i+1,j} - v_{i,j}) + \cos\phi_{i,j}(u_{i,j+1} - u_{i,j}))^2 \\ &+ \sum_{(i,j)\in V} w_{i,j}^2((u_{i+1,j} - u_{i,j}) - \cos\phi_{i,j}(v_{i,j+1} - v_{i,j}))^2 \end{aligned} \tag{2}$$

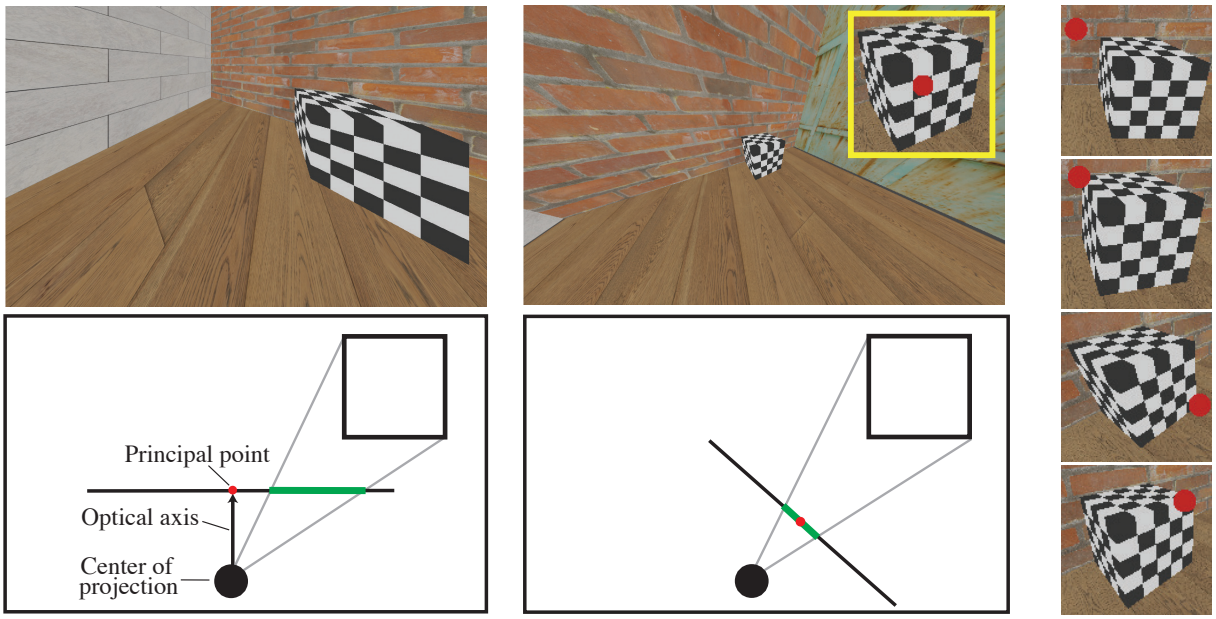
where V denotes the set of boundary vertices (i.e., vertices for which some but not all of $\{(i,j),(i+1,j),(i,j+1)\}$ lie within the region of interest), and $w_{i,j}$ are weights for the energy that are larger in regions with high frequencies and lower in regions with only low frequencies. We use the same weighting w introduced by Carroll et al. [6], but without the term for human faces.

We optimize the objective function E_c with respect to the projection parameters $\{\mathbf{v}_P, a, b, t_x, t_y\}$ using a damped Newton's method with a line search. We initialize $\mathbf{v}_P$ to the direction corresponding to the midpoint between the center of the image and the object mask on the viewing sphere, $a = 1$, and $b = t_x = t_y = 0$.

If multiple regions of interest have been annotated in the input, the same process is repeated separately for each. The transforms T are then input to the next stage.

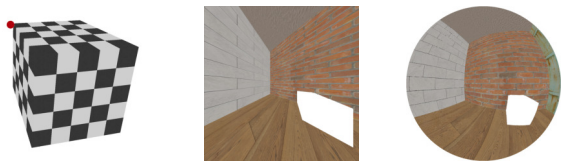
5. Image Warping Algorithm

Given the per-region projections described in the previous Section, MaDCoW optimizes for an entire image projection $f(\mathbf{v})$, parameterized by the per-vertex mapping $\mathbf{p}_{i,j}$. The full distortion correction algorithm is summarized in the supplementary material. The first term in our loss function



(a) Object at image margin (b) Object centered (c) Object reprojected

Figure 4. **Appearance changes under linear perspective.** (a) Linear perspective rendering creates marginal distortion on the cube. (b) Rotating the camera—so that the cube projects to the new principal point—removes the marginal distortion. (c) Using different rotations pointing at the cube, all produce an undistorted cube while subtly changing the appearance, orientation, and scale of the cube. We can parameterize this space of projections by the location of the principal point (marked in red).



(a) Corrected Object (b) Linear perspective (c) Stereographic

Figure 5. **Motivation for stereographic projection.** (a) Our method first reprojects individual regions without distortion. We then need a way to combine this reprojection with the rest of the image. (b) In linear perspective, marginal distortion creates a mismatch between the background and the object shape. (c) Stereographic projection does not produce marginal distortion, and so can match the object boundary much better.

penalizes changes to every vertex contained in any region-of-interest from that region's projection:

$$E_{DVC} = \sum_{k} \sum_{(i,j) \in W_k} ||f(\mathbf{v}_{i,j}) - T_k(\mathbf{v}_{i,j})||^2 \qquad (3)$$

where k indexes regions-of-interest, and T_k is the projection for the k-th mapping, for vertices W_k. The remaining loss terms are drawn from Carroll et al. [6]: preserving straight lines, maximizing conformality, and smoothness. Note that both of these terms can be derived from the DVC: straight lines amount to one-dimensional regions-of-interest, and conformality preserves the DVC for individual quads (Appendix A).

A straight line annotation from $\mathbf{v}_{start}$ to $\mathbf{v}_{end}$ indicates that an arc between these directions on the view sphere maps to a straight line in the output image plane. We sample points $\mathbf{v}_t$ along this arc, where $t \in [0...1]$, and then penalize their distance to the image-space line from $f(\mathbf{v}_{start})$ to $f(\mathbf{v}_{end})$, computed as:

$$E_l = \sum_{t \in [0...1]} ((f(\mathbf{v}_t) - f(\mathbf{v}_{start}))^\top \mathbf{n}_{start,end})^2 \qquad (4)$$

where

$$\mathbf{n}_{start,end} = \mathbf{R}_{90} \frac{f(\mathbf{v}_{end}) - f(\mathbf{v}_{start})}{||f(\mathbf{v}_{end}) - f(\mathbf{v}_{start})||} \qquad (5)$$

where $\mathbf{R}_{90}$ is a 90-degree rotation matrix.

We additionally use a smoothness penalty:

$$E_s = \sum_{(i,j) \in V} w_{i,j}^2 \cos^2 \phi_{i,j} \left\| \begin{bmatrix} u_{i,j+1} - 2u_{i,j} + u_{i,j-1} \\ v_{i,j+1} - 2v_{i,j} + v_{i,j-1} \\ u_{i+1,j+1} - u_{i+1,j} - u_{i,j+1} + u_{i,j} \\ v_{i+1,j+1} - v_{i+1,j} - v_{i,j+1} + v_{i,j} \end{bmatrix} \right\|^2$$

This loss penalizes fast changes in the differential north vector of f, which encourages the warping to be as rigid as possible.

(a) Fewer lines (b) Carroll (c) More lines (d) Carroll

Figure 6. **More lines are not necessarily better.** In this example, we attempt to improve the result of Carroll et al. [6] by adding more line annotations. There are shape distortions in the base of the church in both examples.

The complete loss function is:

$$E_{total} = w_l E_l + w_c E_c + w_s E_s + w_{DVC} E_{DVC} \quad (6)$$

The conformal loss E_c is defined in Equation 2. We optimize this loss with respect to the warp parameters $\mathbf{p}_{i,j}$ using L-BFGS [33] in PyTorch with a learning rate of 1. We set $w_l = 25$, $w_c = 1$, $w_s = 12$, and $w_{DVC} = 1$ for our experiments.

Initialization. For vertices inside any region-of-interest, we interpolate their mappings with their corresponding transform: $\mathbf{p}_{i.j} \leftarrow T(\mathbf{v}_{i,j})$. For vertices outside any region, we interpolate the fitted transforms from the previous step, together with the initial stereographic projection P_{sg}. Let T_k be the transform for the k-th region, and W_k be the set of vertices in the k-th region. We first define the distance transform of a direction on the viewing sphere $\mathbf{v}$ to the nearest direction in region k as:

$$d_k(\mathbf{v}) = \min_{(i,j)\in W_k} ||P_{sg}(\mathbf{v}) - P_{sg}(\mathbf{v}_{i,j})||^2, \quad (7)$$

Then, the blend for all pixels outside of ROIs is:

$$\alpha_k(\mathbf{v}) = \exp(-c\, d_k(\mathbf{v})) \quad (8)$$

$$T'(\mathbf{v}) = \frac{1}{k}\sum_k \alpha_k(\mathbf{v})\, T_k(\mathbf{v}) + (1 - \frac{1}{k}\sum_k \alpha_k(\mathbf{v}))P_{sg}(\mathbf{v})$$

which is applied to all vertices outside of any region: $\mathbf{p}_{i,j} \leftarrow T'(\mathbf{v}_{i,j})$. We set $c = 30$.

6. Results

In this section, we show results using our method to create very wide-angle photographs. In all cases, we captured images using a Ricoh Theta SC2 camera, which captures 360-degree field-of-view images and shows projections from a hemisphere, except for inputs taken from previous work. Figure 7 shows an ablation for the linear projection optimzation, showing that omitting parts of the single-object perspective optimization (Section 4.1) leads to distortions in the output image. Our method takes about two minutes to run on each of the examples and our implementation of Carroll et al. [6] takes about one minute per example. We ablate the effects of the exponential blending, warp initialization, and choice of mesh resolution in the supplementary material.

(a) (b) (c)

Figure 7. **Ablating single-object optimization.** (a) No optimization of the single-object linear perspective, using the principal point at the center of the object. The output has severe distortions. (b) Optimizing the affine transformation with a fixed principal point reduces some, but not all distortions. (c) Our method, which optimizes both, eliminates distortions due to the mismatch between the object and background that appear in the ablations.

Buildings. The state-of-the-art method for panoramic photography authoring is Carroll et al. [6]. Their method takes line annotations as inputs, whereas ours supports both line and ROI annotations. To compare, we replace ROI annotations with line constraints in the same region. For example, in the top row of Figure 1, we use a single ROI constraint for the building on the right. We use a set of line annotations on the building. However, the loss terms for line annotations do not preserve the overall shape of the building, as they compete with the other losses. In Figure 6, we show that simply adding more line constraints does *not* solve the problem. In contrast, our method preserves the overall shape and orientation of the building.

The top two rows of Figure 8 show more where only using straight-line losses produces distortions. In the top row, separate straight-line constraints produces a sheared-looking building, which is unsurprising, since a shear transformation preserves straight lines. In the second row, the building curves toward the horizon.

People. Previous methods recognize faces to correct distortion [6, 26], but leave the rest of the body distorted. The bottom row of Figure 1 compares one of the images provided by Shih et al. [26] as a failure case for their method. In this case, the person on the left has been made wider by

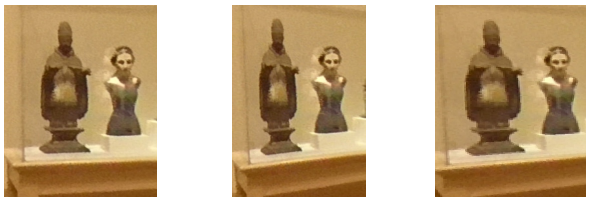
Figure 8. **Results and comparisons.** In each row, the leftmost image shows the results of using only line annotations (Carroll et al. [6]), and the rightmost image shows the result using line and region annotations with MaDCoW. The middle images show zoomed-in views of regions with significant differences, except for the third row, where the middle images show the annotations.

marginal distortion, making him appear particularly overweight; we imagine that few people would like to be made wider in photos [12]. The person on the left now appears to have a normal-width head on a wide body. We treat the face and body together as a single region and correct the person's entire body shape. We show that using Shih et al. [26] with masks over the whole bodies instead of just the faces introduces undesirable curvature in the supplementary material.

The third row of Figure 8 shows four people in a row near the camera. We do not annotate any line constraints here, as the picture has no straight lines; adding them did not help. MaDCoW corrects distortion for the people on the far left and right while keeping their posture upright. In contrast, stereographic projection (i.e., Carroll's method without constraints) corrects the marginal distortion but makes the people appear as if they are slouching toward the center of the image.

Other objects. The fourth row of Figure 8 shows a museum display in which region constraints effectively remove marginal distortion on the sides of the picture; note particularly how the statue on the far left is compressed horizontally when using only line constraints (see Figure 9).

The bottom row of Figure 8 shows the application to a car on the side of a picture. Using only line annotations produces many distortions, e.g., the wheels are not round.



(a) Direct View (b) Carroll et al. [6] (c) MaDCoW

Figure 9. **Close-up of marginal distortion correction.** This statue, on the far right of the fourth row of Figure 8, has considerable marginal distortion. (a) Rendering of the statue at the center of a linear projection. (b) Carroll et al. compresses the object. (c) Our result removes the marginal distortion.

	1	2	3	4	5	6	7
Object	95%	84%	93%	79%	95%	75%	65%
Overall	95%	68%	90%	75%	90%	84%	59%

Table 1. User study results, comparing to a version with line annotations only, as in [6]; except for column 2, which shows a comparison to Shih et al.'s result [26]). Participants were asked to compare objects in the margins (first row), and to evaluate the entire output images (second row). Columns are ordered by the order in which they appear in this paper, i.e., column 1 is the first row of Figure 1, column 3 is the first row of Figure 8, etc. In all cases, viewers preferred the results of our method.

User study. We performed a user study in which viewers were shown pairs of image results from this paper, paired with Carroll's results, and asked two questions: which results had more natural-looking objects on the margins when compared to a direct view, and which results looked better as a whole, to 44 participants. Overall, viewers preferred our result 81% of the time and thought our results had more natural-looking objects 84% of the time; see Table 1.

Limitations. Our results require the user to annotate the target distorted regions, e.g., masking a building. Our optimization works best when ROIs have sufficient distance between them to allow gradual transitions—correcting distortions on densely packed regions may produce undesirable artifacts. Because image features connect to one another, e.g., a sidewalk that abuts the building, the optimal mask may be non-obvious and need user refinement. Salient image features that cross multiple ROIs, e.g., a power line that goes from the sky to in front of a building, and will appear bent at the mask boundary; regions that span large fields-of-view will have some marginal distortion regardless, e.g., a large building with straight lines that span the picture.

7. Conclusion

We present MaDCoW, which can produce effective wide-angle panoramas without the distortions produced by previous methods. We achieve this by optimizing a local linear projection for each ROI, then a whole-image optimization combining all constraints, but we anticipate that this general strategy can be expanded to other capture scenarios and further improved upon.

MaDCoW relies on an optimization that is not fast enough for real-time user control. One approach to faster performance would be to use neural networks to directly predict the warp in a feedfoward manner, as in Zhu et al. [34].

Like previous methods, MaDCoW uses user input, as the choice of projection depends on the user's goals. Further research could help determine common goals in creating wide-angle images, perhaps by studying how users use tools for authoring wide-angle images and what kinds of wide-angle images they prefer, or by taking inspiration from art history [13, 15]. Since different users may have different goals, future work could build systems that let people easily capture and author wide-angle projections, for example, a one-click selection of objects to emphasize. One could also imagine a fully automatic system that detects classes of objects, similar to Zhang et al. [32], and applies appropriate projections to them.

A. Conformality satisfies DVC

Here we sketch out a proof that satisfying conformality satisfies the DVC.

Each vertex (i, j) in the input mesh corresponds to an input direction on the view sphere $\mathbf{v} = (\lambda, \phi)$ centered at the center-of-projection (Figure 3). A triangle on the input mesh is given by indices $(i, j), (i+1, j), (i, j+1)$. We can define an image plane as the plane containing this triangle; the set of rays through this plane to the center-of-projection is a linear perspective projection $P(\mathbf{v})$. The principal point for this projection is given by the closest point from the plane, and this point is guaranteed to lie in or near the triangle. Hence, this projection satisfies the DVC.

We map each vertex (i, j) to image coordinates $\mathbf{p}_{i,j} = (u_{i,j}, v_{i,j})$; each view sphere triangle maps to an image space triangle with vertices $\mathbf{p}_{i,j}, \mathbf{p}_{i+1,j}, \mathbf{p}_{i,j+1}$. This defines a new projection $T(\mathbf{v})$ as the original projection $P(\mathbf{v})$ followed by an affine transform in image space. The mapping satisfies the DVC if and only if the affine component includes only rigid and uniform scale transformations and no shear or nonuniform stretch. This is enforced by the discrete conformality conditions from Carroll et al. [6] (Equation 2). That is, the conformality loss is nonzero if and only if there are shear or nonuniform stretches, which amount to scaling u and v differently. In short, enforcing conformality enforces the DVC.

Taking the limit of infinitesimal triangles gives the analogous continuous result, i.e., defined on the partial derivatives of f.

References

[1] Aseem Agarwala, Maneesh Agrawala, Michael Cohen, David Salesin, and Richard Szeliski. Photographing long scenes with multi-viewpoint panoramas. *ACM Trans. Graphics*, 25(3):853–861, 2006. 2

[2] Maneesh Agrawala, Denis Zorin, and Tamara Munzner. Artistic multiprojection rendering. In *Eurographics Workshop on Rendering Techniques*, pages 125–136. Springer, 2000. 2

[3] Hadi Alzayer, Zhihao Xia, Xuaner Zhang, Eli Shechtman, Jia-Bin Huang, and Michael Gharbi. Magic fixup: Streamlining photo editing by watching dynamic videos, 2024. 2

[4] Abhishek Badki, Orazio Gallo, Jan Kautz, and Pradeep Sen. Computational zoom: A framework for post-capture image composition. *ACM Trans. Graph.*, 36(4), 2017. 3

[5] Alistair Burleigh, Robert Pepperell, and Nicole Ruta. Natural perspective: Mapping visual space with art and science. *Vision*, 2(2):21, 2018. 3

[6] Robert Carroll, Maneesh Agrawala, and Aseem Agarwala. Optimizing content-preserving projections for wide-angle images. *ACM Trans. Graph.*, 28(3), 2009. 1, 2, 3, 4, 5, 6, 7, 8

[7] Robert Carroll, Aseem Agarwala, and Maneesh Agrawala. Image warps for artistic perspective manipulation. *ACM Trans. Graph.*, 29(4), 2010. 2

[8] Che-Han Chang, Min-Chun Hu, Wen-Huang Cheng, and Yung-Yu Chuang. Rectangling stereographic projection for wide-angle image visualization. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2013. 3

[9] Ilya Chugunov, Amogh Joshi, Kiran Murthy, Francois Bleibel, and Felix Heide. Neural light spheres for Implicit Image Stitching and View Synthesis. In *Proceedings of the ACM SIGGRAPH Asia 2024*. ACM, 2024. 2

[10] Emily A Cooper, Elise A Piazza, and Martin S Banks. The perceptual basis of common photographic practice. *Journal of vision*, 12(5):8–8, 2012. 1

[11] Ohad Fried, Eli Shechtman, Dan B Goldman, and Adam Finkelstein. Perspective-aware manipulation of portrait photos. *ACM Transactions on Graphics (TOG)*, 35(4):1–10, 2016. 3

[12] Ohad Fried, Jennifer Jacobs, Adam Finkelstein, and Maneesh Agrawala. Editing self-image. *Communications of the ACM*, 63(3):70–79, 2020. 7

[13] Aaron Hertzmann. Toward a theory of perspective perception in pictures. *Journal of Vision*, 24(4), 2024. 1, 2, 4, 8

[14] Liao Kang, Yue Zongsheng, Wu Zhonghua, and Chen Change Loy. Mowa: Multiple-in-one image warping model. *arXiv:2404.10716*, 2024. 3

[15] Jan Koenderink, Andrea van Doorn, Baingio Pinna, and Robert Pepperell. On right and wrong drawings. *Art and Perception*, 4(1-2):1–38, 2016. 2, 4, 8

[16] Michael Kubovy. *The Psychology of Perspective and Renaissance Art*. Cambridge University Press, 1986. 2

[17] Wei-Sheng Lai, YiChang Shih, Chia-Kai Liang, and Ming-Hsuan Yang. Correcting face distortion in wide-angle videos. *IEEE Transactions on Image Processing*, 2021. 2

[18] Hyunjoon Lee, Eli Shechtman, Jue Wang, and Seungyong Lee. Automatic upright adjustment of photographs with robust camera calibration. *IEEE transactions on pattern analysis and machine intelligence*, 36(5):833–844, 2013. 2, 3

[19] Kang Liao, Chunyu Lin, Yao Zhao, and Moncef Gabbouj. DR-GAN: Automatic Radial Distortion Rectification Using Conditional GAN in Real-Time. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(3):725–733, 2020. 3

[20] Kang Liao, Chunyu Lin, Yao Zhao, and Mai Xu. Model-Free Distortion Rectification Framework Bridged by Distortion Distribution Map. *IEEE Transactions on Image Processing*, 29:3707–3718, 2020. 3

[21] Sean J Liu, Maneesh Agrawala, Stephen DiVerdi, and Aaron Hertzmann. Zoomshop: Depth-aware editing of photographic composition. *Computer Graphics Forum*, 41(2):57–70, 2022. 3

[22] Manuel López, Roger Marí, Pau Gargallo, Yubin Kuang, Javier Gonzalez-Jimenez, and Gloria Haro. Deep Single Image Camera Calibration With Radial Distortion. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11809–11817, 2019. 3

[23] Augusto Román and Hendrik PA Lensch. Automatic multiperspective images. *Rendering Techniques*, 2(2006):161–171, 2006. 2

[24] Augusto Román, Gaurav Garg, and Marc Levoy. Interactive design of multi-perspective images for visualizing urban landscapes. In *IEEE visualization 2004*, pages 537–544. IEEE, 2004. 2

[25] Thomas K Sharpless, Bruno Postle, and Daniel M German. Pannini: A new projection for rendering wide angle perspective images. In *Computational Aesthetics*, pages 9–16, 2010. 3

[26] YiChang Shih, Wei-Sheng Lai, and Chia-Kai Liang. Distortion-free wide-angle portraits on camera phones. *ACM Trans. Graph.*, 38(4), 2019. 1, 2, 4, 6, 7, 8

[27] Richard Szeliski. Image alignment and stitching: A tutorial. *Foundations and Trends in Computer Graphics and Vision*, 2 (1), 2007. 2

[28] Mahdi Abbaspour Tehrani, Aditi 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. 2

[29] Jin Wei, Chen-Feng Li, Shi-Min Hu, Ralph R. Martin, and Chiew-Lan Tai. Fisheye Video Correction. *IEEE Transactions on Visualization and Computer Graphics*, 18(10): 1771–1783, 2012. 3

[30] Jingyi Yu, Leonard McMillan, and Peter Sturm. Multiperspective Modeling, Rendering, and Imaging. In *Eurographics 2008 - State of the Art Reports*. The Eurographics Association, 2008. 3

[31] Lihi Zelnik-Manor, Gabriele Peters, and Pietro Perona. Squaring the circle in panoramas. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, pages 1292–1299. IEEE, 2005. 2, 4

[32] Qi Zhang, Hongdong Li, and Qing Wang. Wide-angle rectification via content-aware conformal mapping. In *CVPR*, 2023. 3, 8

[33] Ciyou Zhu, Richard H Byrd, Peihuang Lu, and Jorge Nocedal. Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on mathematical software (TOMS)*, 23(4):550–560, 1997. 6

[34] Fushun Zhu, Shan Zhao, Peng Wang, Hao Wang, Hua Yan, and Shuaicheng Liu. Semi-supervised wide-angle portraits correction by multi-scale transformer. In *CVPR*, 2022. 2, 8

[35] Denis Zorin and Alan 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, 4