

Squaring the Circle in Panoramas

Lihi Zelnik-Manor¹

Gabriele Peters²

Pietro Perona¹

1. Department of Electrical Engineering
California Institute of Technology
Pasadena, CA 91125, USA

2. Informatik VII (Graphische Systeme)
Universitat Dortmund
Dortmund, Germany

<http://www.vision.caltech.edu/lihi/SquarePanorama.html>

Abstract

Pictures taken by a rotating camera cover the viewing sphere surrounding the center of rotation. Having a set of images registered and blended on the sphere what is left to be done, in order to obtain a flat panorama, is projecting the spherical image onto a picture plane. This step is unfortunately not obvious – the surface of the sphere may not be flattened onto a page without some form of distortion. The objective of this paper is discussing the difficulties and opportunities that are connected to the projection from viewing sphere to image plane. We first explore a number of alternatives to the commonly used linear perspective projection. These are ‘global’ projections and do not depend on image content. We then show that multiple projections may coexist successfully in the same mosaic: these projections are chosen locally and depend on what is present in the pictures. We show that such multi-view projections can produce more compelling results than the global projections.

1. Introduction

As we explore a scene we turn our eyes and head and capture images in a wide field of view. Recent advances in storage, computation and display technology have made it possible to develop ‘virtual reality’ environments where the user feels ‘immersed’ in a virtual scene and can explore it by moving within it. However, the humble still picture, painted or printed on a flat surface, is still a popular medium: it is inexpensive to reproduce, easy and convenient to carry, store and display. Even more importantly, it has unrivaled size, resolution and contrast. Furthermore, the advent of inexpensive digital cameras, their seamless integration with computers, and recent progress in detecting and matching informative image features [4] together with the development of good blending techniques [7, 5] have made it possible for any amateur photographer to produce automatically mosaics of photographs covering very wide fields of view

and conveying the vivid visual impression of large panoramas. Such mosaics are superior to panoramic pictures taken with conventional fish-eye lenses in many respects: they may span wider fields of view, they have unlimited resolution, they make use of cheaper optics and they are not restricted to the projection geometry imposed by the lens.

The geometry of single view point panoramas has long been well understood [12, 21]. This has been used for mosaicing of video sequences (e.g., [13, 20]) as well as for obtaining super-resolution images (e.g., [6, 23]). By contrast when the point of view changes the mosaic is ‘impossible’ unless the structure of the scene is very special. When all pictures share the same center of projection C , we can consider the viewing sphere, i.e., the unit sphere centered in C , and identify each pixel in each picture with the ray connecting C with that pixel and passing through the surface of the viewing sphere, as well as through the physical point in the scene that is imaged by that pixel. By detecting and matching visual features in different images we may register automatically the images with respect to each other. We may then map every pixel of every image we collected to the corresponding point of the viewing sphere and obtain a spherical image that summarizes all our information on the scene. This spherical image is the most natural representation: we may represent this way a scene of arbitrary angular width and if we place our head in C , the center of the sphere, we may rotate it around and capture the same images as if we were in the scene.

What is left to be done, in order to obtain our panorama-on-a-page, is projecting the spherical image onto a picture plane. This step is unfortunately not obvious – the surface of the sphere may not be flattened onto a page without some form of distortion. The choice of projection from the sphere to the plane has been dealt with extensively by painters and cartographers. An excellent review is provided in [9].

The best known projection is **linear perspective** (also called ‘gnomonic’ and ‘rectilinear’). It may be obtained by projecting the relevant points of the viewing sphere onto a tangent plane, by means of rays emanating from the cen-

ter of the sphere C . Linear perspective became popular amongst painters during the Renaissance. Brunelleschi is credited with being the first to use correct linear perspective. Alberti wrote the first textbook on linear perspective describing the main construction methods [1]. It is believed by many to be the only ‘correct’ projection because it maps lines in 3D space to lines on the 2D image plane and because when the picture is viewed from one special point, the ‘center of projection’ of the picture, the retinal image that is obtained is the same as when observing the original scene. A further, somewhat unexpected, virtue is that perspective pictures look ‘correct’ even if the viewer moves away from the center of projection, a very useful phenomenon called ‘robustness of perspective’ [18, 22].

Unfortunately, linear perspective has a number of drawbacks. First of all: it may only represent scenes that are at most 180° wide: as the field of view becomes wider, the area of the tangent plane dedicated to representing one degree of visual angle in the peripheral portion of the picture becomes very large compared to the center, and eventually becomes unbounded. Second, there is an even more stringent limit to the size of the visual field that may be represented successfully using linear perspective: beyond widths of 30° - 40° architectural structures (parallelepipeds and cylinders) appear to be distorted, despite the fact that their edges are straight [18, 14]. Furthermore, spheres that are not in the center of the viewing field project to ellipses onto the image plane and appear unnatural and distorted [18] (see Fig 1). Renaissance painters knew of these shortcomings and adopted a number of corrective measures [14], some of which we will discuss later.

The objective of this paper is discussing the difficulties and opportunities that are connected to the projection from viewing sphere to image plane, in the context of digital image mosaics. We first argue that linear perspective is not the only valid projections, and explore a number of alternatives which were developed by artists and cartographers. These are ‘global’ projections and do not depend on image content. We explore experimentally the tradeoffs of these projections: how they distort architecture and people and how well do they tolerate wide fields of view. We then show that multiple projections may coexist successfully in the same mosaic: these projections are chosen locally and depend on what is seen in the pictures that form the mosaic. We show experimentally that the such a content based approach often produces more compelling results than the traditional approach which utilizes geometry only. We conclude with a discussion of the work that lies ahead.

In this paper we do not address issues of image registration and image blending. In all our experiments the computation of the transformations of the input images to the sphere and the blending of the images was done using Matthew Brown’s Autostitch software [4, 2].

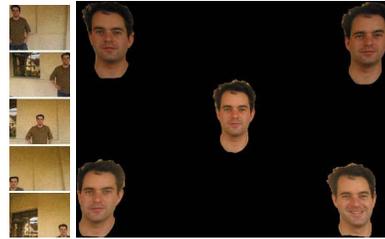


Figure 1. Perspective distortions. *Left: Five photographs of the same person taken by a rotating camera, after rectification (removing spherical lens distortion). Right: An overlay of the five photographs after blackening everything but the person’s face shows that spherical objects look distorted under perspective projection even at mild viewing angles. For example, here, the centers of the faces in the corners are at $\sim 20^\circ$ horizontal eccentricity.*

2 Global Projections

What are the alternatives to linear perspective?

An important drawback of linear perspective is the excessive scaling of sizes at high eccentricities. Consider a painter taking measurements in the scene by using her thumb and using these measurements to scale objects on the canvas. She takes angular measurements in the scene and translates them into linear measurements onto the canvas. This construction is called **Postel** projection [9]. It avoids the ‘explosion’ of sizes in the periphery of the picture. Along lines radiating from the point where the picture plane touches the viewing sphere, it actually maps lengths on the sphere to equal lengths in the image. Lines that run orthogonal to those (i.e., concentric circles around the tangent point) will be magnified at higher eccentricities, but much less than by linear perspective. The Postel projection is close to the cartographic **stereographic** projection. The stereographic projection is obtained by using the pole opposite to the point of tangency as the center of projection.

Consider now the situation in which we wish to represent a very wide field of view. A viewer contemplating a wide panorama will rotate his head around a vertical axis in order to take in the full view. Suppose now that the view has been transformed into a flat picture hanging on a wall and consider a viewer exploring that picture: the viewer will walk in front of the picture with a translatory motion that is parallel to the wall. If we replace rotation around a vertical axis with sideways translation in front of the picture we obtain a family of projections which are popular with cartographers. Wrap a sheet of paper around the viewing sphere forming a cylinder that touches the sphere at the equator. One may project the meridians onto the cylinder by maintaining lengths along vertical lines, thus obtaining the **ge-**

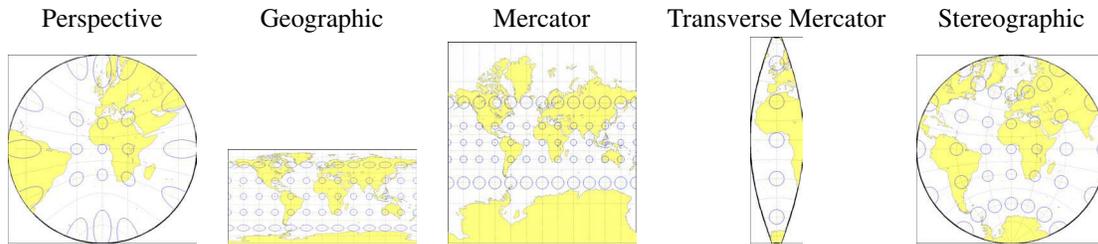


Figure 2. Spherical projections. Figures taken out of Matlab’s help pages visualizing the distortions of various projections. Grid lines correspond to longitude and latitude lines. Small circles are placed at regular intervals across the globe. After projection, the small circles appear as ellipses (called Tissot indicatrices) of various sizes, elongations, and orientations. The sizes and shapes of the ellipses reflect the projection distortions.

ographic projection. Alternatively, one may want to vary locally the scale of the meridians so that they keep in proportion with the parallels. This is the **Mercator** projection (for mathematical definitions of these projections see [16]).

Figure 2 visualizes the properties of these projections. Grid lines correspond to longitude and latitude lines. When projecting images onto the sphere, vertical lines are projected onto longitude lines. Horizontal lines are not projected onto latitude lines but rather onto tilted great circles, thus the visualization of the latitude lines does not convey what happens to horizontal image lines. All of these projections are global and are independent of the image content.

Figure 3 illustrates the above projections on a panorama constructed of images taken at an indoor scene. This is a typical example of panoramas of man-made environments which usually contain many straight lines. Selecting from the above projections implies bending either the horizontal lines, the vertical lines, or both. In most cases a better choice is to keep vertical lines straight as this results in a panorama where narrow vertical slits look correct. This matches the observations in [22], which shows that our perception of a picture is affected by the fact that normally people shift their gaze horizontally and rarely shift it vertically. Shifting one’s gaze horizontally across a panorama looks best when vertical lines are not bent. This motivates the use of either the Geographic or the Mercator projections. In both these projections the rotation of the camera is transformed into sideways motion of the observer.

When the camera performs mostly pan motion, i.e., when the vertical angle is small, both projections produce practically the same result. However, for larger tilt angles the Geographic projection distorts circles, i.e., it does not maintain correct proportions, while the Mercator does maintain conformality, thus the Mercator projection is a better option (see Figure 4). Note, that the conformality implies that in the Mercator projection spherical and cylindrical objects, such as people, are not distorted but the background is, see for example Figure 7.

An important issue in all cylindrical projections is the

choice of equator. Once the images are on the sphere one can rotate the sphere in any desired way before projecting to the plane. In other words, the cylinder wrapping the sphere can touch the sphere along an equator of choice. When a wrong equator is selected, vertical lines in 3D space will not be projected onto vertical lines in the panorama (see left panel of Figure 5). Finding the correct equator is easy. The user is requested to mark a single vertical line and a horizon point in one (or two) of the input images. The sphere is then rotated so that projection of the marked vertical line aligns with a longitude line and the equator goes through the selected horizon point. This results in a straightened panorama, see for example, right panel of Figure 5.

Should other projections be considered? Yes, we think so. The **Transverse Mercator** projection is known in the mapping world as an excellent choice for mapping areas that are elongated north-to-south. This corresponds to panoramas with little pan motion and large tilt motion. The bending of vertical lines is small near the meridian, thus, when the pan angle is small we are better off using the Transverse Mercator projection which keeps the horizontal lines straight. This is illustrated in Figures 4, 6.

For far away outdoors scenes almost any projection looks good as the scenes rarely contain any straight lines. Nevertheless, too much bending might disturb the eye even on free form objects like clouds. This implies the usage of the stereographic projection, which bends both vertical and horizontal lines but less than the cylindrical projections.

So, which is the best option? Unfortunately, there is no single answer. As discussed in above, each photographing setup will have its own best solution. Instead, in the next section we suggest a multi-view approach and show experimentally that it produces better looking mosaics.

3 Multi View Projection

The projections explored in Section 2 are ‘global’, in that once a tangent point or a tangent line is chosen, the projection is completely determined by this parameter. As can be

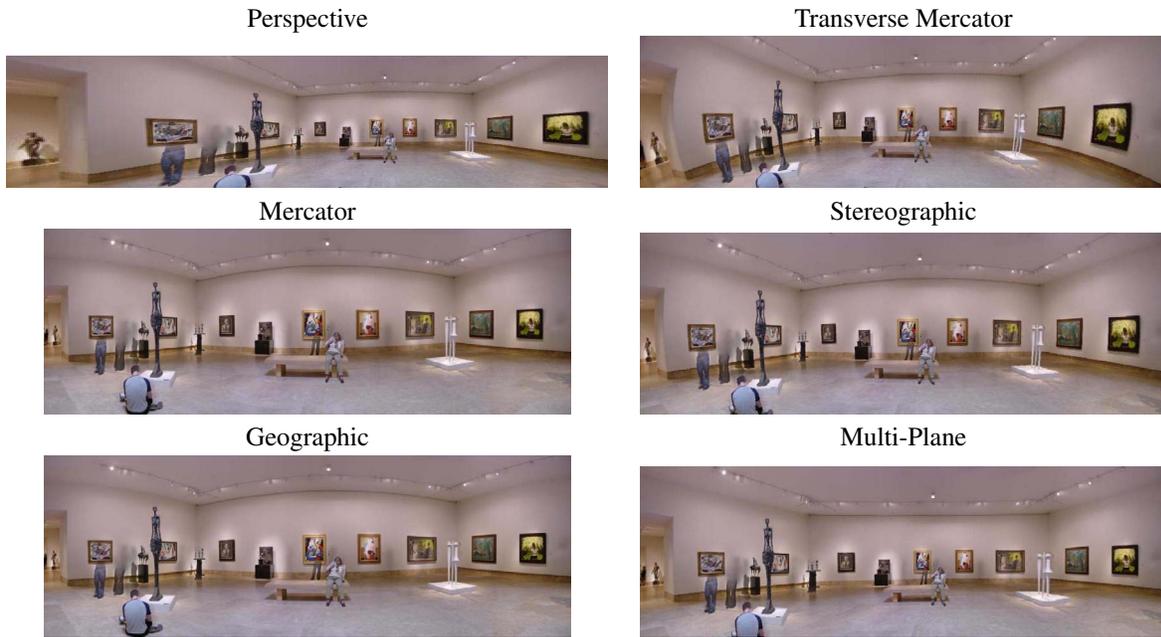


Figure 3. Spherical projections. *There are many spherical projections. Each has its pros and cons.*

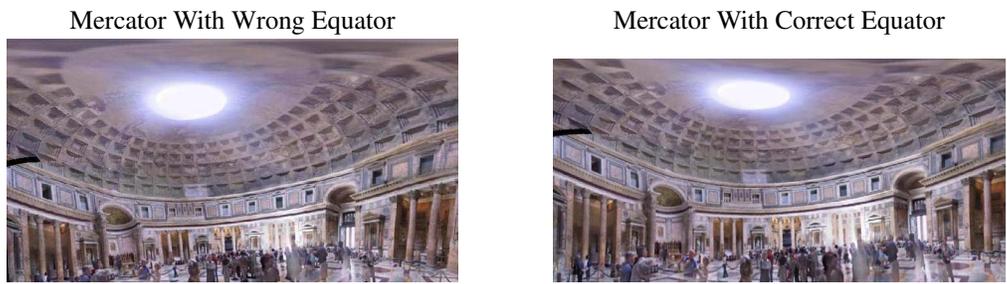


Figure 5. Choice of equator (*The Pantheon*). *A wrong choice of the equator results in tilted vertical lines. The columns on the right and left appear converging. Correcting the equator selection results in columns standing up-right.*

seen in the example figures, they all introduce some distortions. Zorin and Barr [24] suggested an approach to remove distortions from perspective images by compromising between bending lines and distorting spheres. They present impressive results, however, their solution is limited to fields of view which can be covered by the linear perspective projection. Using a single global projection is by no means a necessary property for a good projection. We may instead tailor the projection locally to the content of the images in order to improve the final effect. We next explore a few options for such multi-view projections.

3.1 Multi-Plane Perspective Projection

As was shown in Section 2, a global projection of wide panoramas bends lines, which is unpleasant to the eye. To obtain both a rectilinear appearance and a large field of view

we suggest using a multi-plane perspective projection. Such multi-plane projections were suggested by Greene [11] for rendering textured surfaces. Rather than projecting the sphere onto a single plane, multiple tangent planes to the sphere are used. Each projection is linear perspective. The tangent planes have to be arranged so that they may be unfolded into a flat surface without distortion, e.g., the points of tangency belong to a maximal circle. One may think of the intersections of the tangent planes being fitted with hinges that allow flattening. The projection onto each plane is perspective and covers only a limited field of view, thus it is pleasant to the eye.

This process introduces large orientation discontinuities at the intersection between the projection planes, however, in many man-made environment these discontinuities will not be noticed if they occur along natural discontinuities. The tangent planes must therefore be chosen in a way that

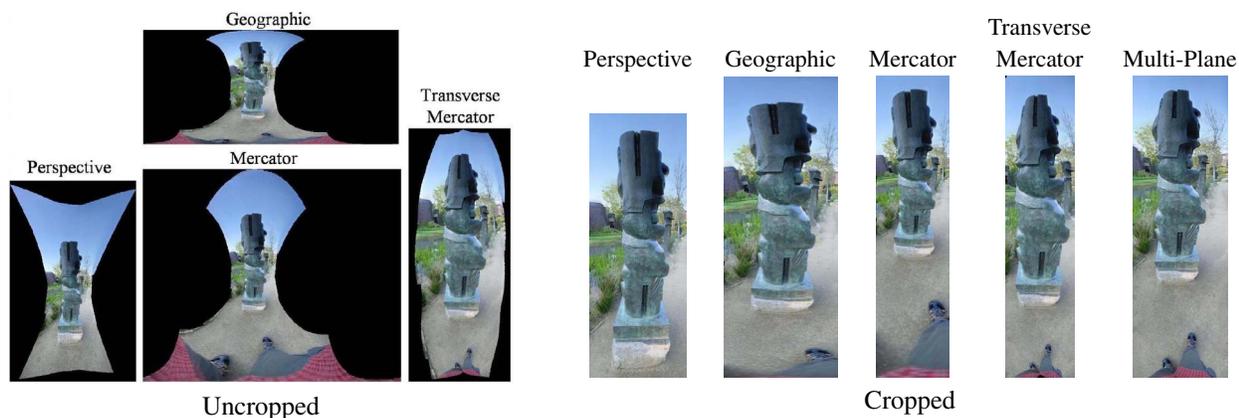


Figure 6. Vertical panoramas. *Left and right panels show results before and after cropping (see Section 3.3 for further details). For wide angle panoramas, perspective cannot capture the full range, thus the photographers legs are excluded. Geographic distorts proportions (see how squashed the legs look). Mercator stretches the legs across the bottom. Transverse-Mercator captures both the sculpture and the photographer which suggests it is the best global projection option for narrow vertical panoramas. Multi-Plane does even better.*

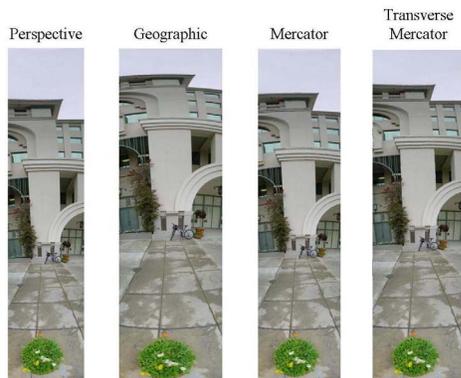


Figure 4. Preserving proportions. *The Geographic projection distorts the circular pot at the bottom into an ellipse. The Mercator projection does not.*

fits the geometry of the scene, e.g. so that the vertical edges of a room project onto the seams and each projection plane corresponds to a single wall. Orientation discontinuities caused by the projection this way co-occur with orientation discontinuities in the scene and therefore they are visually unnoticeable (see Figures 3, 7, 6). Sometimes no seam may be found that completely corresponds to discontinuities in the scene: for example in Figure 9 the chair on the right is clearly distorted. Another caveat is that some arrangements will cause a loss in the impression of depth: for example, when projecting a panorama of a standard room onto a square prism (see left panel of Figure 8). Most often the sensation of depth can be maintained by appropriate choice of the projection planes (see right panel of figure 8).

We have currently implemented a simple user interface to allow choosing the position of the multiple tan-

gent planes. We assume that the hinges between tangent planes are either associated to vertical or horizontal lines: the user is presented with the Geographic projection of the panorama and clicks once anywhere on a single vertical line to choose a seam and once again to choose the point of tangency of each projection plane. Automating this operation is an interesting exercise which we leave for the future.

3.2 Preserving Foreground Objects

The multi-plane perspective projection takes us back to the second challenge presented in Section 1. Recall, that even for small fields of view nearby (foreground) objects are often perceived as distorted. Our solution to this problem draws its inspiration from the Renaissance artists.

During the Renaissance the rules of perspective were understood, and linear perspective was used to produce pictures that had a realistic look. Painters noticed earlier on, that spheres and cylinders (and therefore people) would appear distorted if they were painted according to the rules of a global perspective projection (a sphere will project to an ellipse). It thus became common practice to paint people, spheres and cylinders by using linear perspective centered around each object. (see for example the *The School of Athens* by Raphael [18, 14]). This results in paintings with multiple view points. There is one global view point used for the background and an additional view point for each foreground person/object.

Renaissance paintings look good precisely because they are constructed using a multiplicity of projections. Each projection is chosen in order to minimize the apparent distortion of either the ambient architecture, or of a specific person/object. We follow this example and adopt the multi-

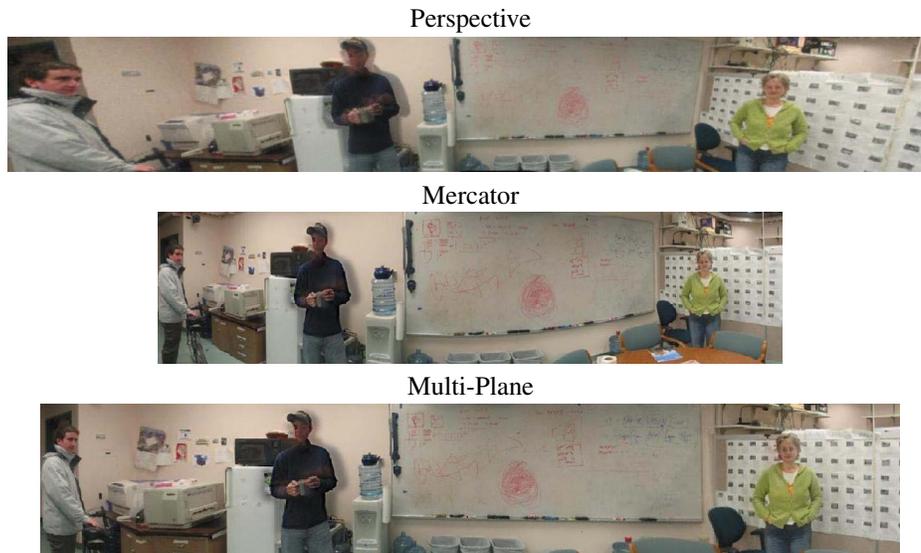


Figure 7. Architecture vs. spherical objects. *Perspective projection distorts people at large viewing angles. Mercator projection keeps the people undistorted, but distorts the wall and white-board at the background. The Multi-Plane projection provides the most compelling result with no noticeable distortions in both background and people.*

view point approach to construct realistic looking panoramas. We first separate the background and foreground objects. A panorama is constructed from the background by using a global projection: perspective for fields of view that are narrower than, say, 40° and Multi-Plane otherwise. The foreground objects are projected using a ‘local’ perspective projection, with a central line of sight going through the center of each object, and then they are pasted onto the background. More in detail:

- (1) Obtain a foreground-background segmentation for each image and cut out the foreground objects [15, 19]. Currently we use the GIMP [10] implementation of Intelligent Scissors [17] which requires manual interaction, we found it to take less than a minute per image.
- (2) Fill in the holes in the background caused by cutting out the foreground objects using a texture propagation technique (e.g., [8, 3]). We used our implementation of [8]. Note, that the hole filling need not be perfect as most of it will be covered eventually by the repasting of the foreground objects. As we are most sensitive to people’s distortions, one could acquire each picture containing a person a second time, once the person moved. In that case hole filling won’t be required.
- (3) Build a panorama of the filled background images.
- (4) Overlay foreground objects on top of the background panorama. For each foreground object, find its bounding box in the original image and in the panorama if it were projected along with the background. Rescale the cut-out object to have the same height as its projection (note, that the width will be different). Paste the object so that the centers of the bounding boxes align.

This process is illustrated in Figure 10. Five frames were taken out of a video sequence showing a child walking from right to left, while facing the camera. The child was cut-out from each image, texture propagation was used to fill in the holes and a perspective panorama of the background was constructed (see Figure 10 top). The cut-outs of the child were then pasted onto the background in two ways. Once applying the same perspective projection used for the background, which resulted in distorting the child’s head into a variety of ellipsoidal shapes (see Figure 10 middle). Then using the multi-view approach described above which produced a significantly better looking result, removing all the head distortions, see Figure 10 bottom. Another example is displayed in Figure 11 (for this example we had available clear background images so hole filling was not required). Figure 9 displays our full solution including both multi-plane projection for the background and multi-view projection to correct the chair in the foreground.

3.3 Visualization Details

When the images do not cover the full viewing sphere, the boundaries of the panorama can have all sort of shapes, depending on the projection, e.g., see left panel of Figure 6. Thus, for visualization purposes, the panoramas were cropped to display a complete rectangular portion. This results in different coverage areas for each projection. The uncropped panoramas, as well as more results can be found at: <http://www.vision.caltech.edu/lihi/SquarePanorama.html>

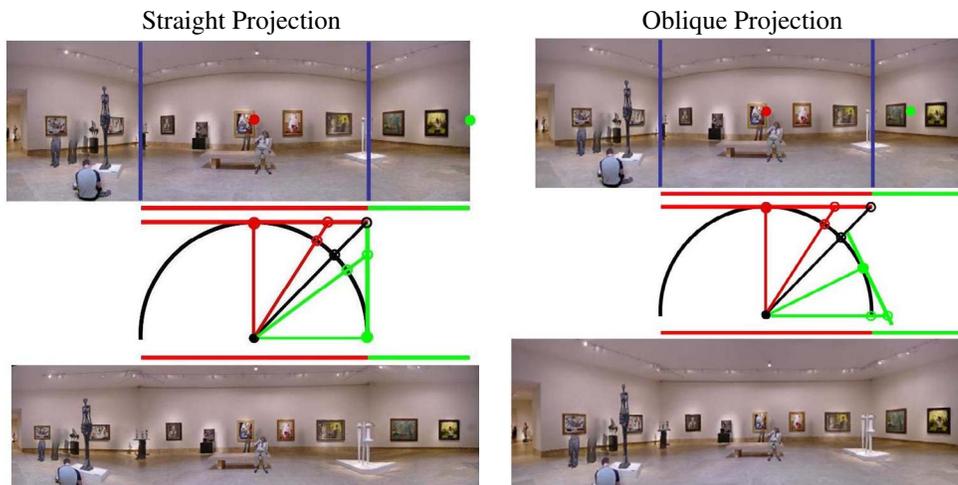


Figure 8. Multi-Plane projection. In each panel the top figure displays the geographic projection and the interaction required by the user - definition of the intersection lines between the tangent planes (marked in blue) and the center of projection for each tangent plane (marked in green and red). The middle panel displays a top view of the projection. The bottom panel displays the final result.



Figure 9. Multi-Plane Multi-View. The multi-plane projection rectified the background but the chair on the right got distorted. Using the Multi-View approach the chair is undistorted.

4 Discussion & Conclusions

The challenge of constructing panoramas from images taken from a single viewpoint goes beyond image matching and image blending. The choice of the mapping between the viewing sphere and the image plane is an interesting problem in itself. Artists and cartographers have explored this problem and have proposed a number of useful global projections. Additionally, artists have developed a practice to use multiple local projections which are guided by the content of the images. Inspired by the artists we have proposed a new set of projections which incorporate multiple local projections with multiple view points into the same panorama to produce more compelling results. Further automating this process is a worthwhile challenge for machine vision researchers.

5 Acknowledgements

This research was supported by MURI award number AS3318 and the Center of Neuromorphic Systems Engineering award EEC-9402726. We also wish to acknowledge our useful conversations with Pat Hanrahan, Jan Koenckerink, Marty Banks, Bill Freeman, Ged Ridgway and David Lowe and to thank Matthew Brown for providing his Autostitch software.

References

- [1] L. B. Alberti. *On Painting*. First appeared 1435-36. Translated with Introduction and Notes by John R. Spencer. New Haven: Yale University Press, 1970.
- [2] Autostitch. <http://www.autostitch.net/>.
- [3] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. In *Proceedings of SIGGRAPH*, New Orleans, USA, July 2000.

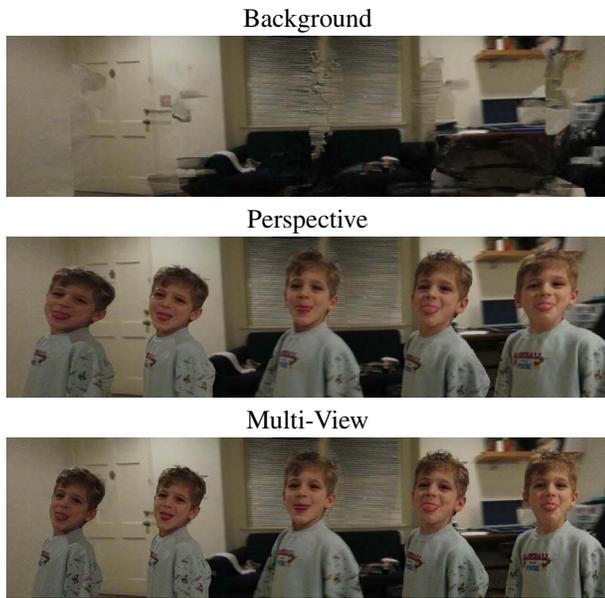


Figure 10. Correcting perspective distortions. Top: Panorama of the background only. Artifacts in the hole filling are visible, but are inessential as they will be eventually covered by the foreground object. Center: A global perspective projection of both background and foreground. The child's head appears distorted. Bottom: A multi-view point panorama providing the most compelling look with no head distortions.



Figure 11. Correcting perspective distortions. In the Perspective panorama the person's head is highly distorted. A Multi-view panorama provides a more compelling look, removing all distortions.

[4] M. Brown and D. Lowe. Recognising panoramas. In *Proceedings of the 9th International Conference on Computer Vision*, volume 2, pages 1218–1225, Nice, October 2003.

[5] P. J. Burt and E. H. Adelson. A multiresolution spline with application to image mosaics. *ACM Trans. Graph.*, 2(4):217–236, 1983.

[6] D. Capel and A. Zisserman. Automatic mosaicing with super-resolution zoom. In *CVPR '98: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, page 885. IEEE Computer Society, 1998.

[7] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *Proceedings of SIGGRAPH*, August 1997.

[8] A. Efros and T. K. Leung. Texture synthesis by non-parametric sampling. In *IEEE International Conference on Computer Vision*, pages 1033–1038, Corfu, Greece, September 1999.

[9] A. Flocon and A. Barre. *Curvilinear Perspective, From Visual Space to the Constructed Image*. University of California Press, 1987.

[10] T. GIMP. <http://www.gimp.org/>.

[11] N. Greene. Environment mapping and other applications of world projections. *IEEE Computer Graphics and Applications*, 6(11):21–29, November 1986.

[12] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[13] M. Irani, B. Rousso, and S. Peleg. Computing occluding and transparent motions. *Int. J. Comput. Vision*, 12(1):5–16, 1994.

[14] M. Kubovy. *The Psychology of Perspective and Renaissance Art*. Cambridge University Press, 1986.

[15] Y. Li, J. Sun, C. Tang, and H. Shum. Lazy snapping. In *Proceedings of SIGGRAPH*, 2004.

[16] MathWorld. <http://mathworld.wolfram.com/>.

[17] E. Mortensen and W. Barrett. Intelligent scissors for image composition. In *SIGGRAPH '95: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 191–198. ACM Press, 1995.

[18] M. H. Pirenne. *Optics, Painting & Photography*. Cambridge University Press, 1970.

[19] C. Rother, V. Kolmogorov, and A. Blake. Grabcut - interactive foreground extraction using iterated graph cuts. *Proc. ACM Siggraph*, 2004.

[20] H. S. Sawhney and R. Kumar. True multi-image alignment and its application to mosaicing and lens distortion correction. *IEEE Trans. Pattern Anal. Mach. Intell.*, 21(3):235–243, 1999.

[21] R. Szeliski and H. Shum. Creating full view panoramic image mosaics and environment maps. *Computer Graphics*, 31(Annual Conference Series):251–258, 1997.

[22] D. Vishwanath, A. R. Girshick, and M. S. Banks. Why pictures look right when viewed from the wrong place. Personal communication. (Manuscript accepted for publication).

[23] A. Zomet and S. Peleg. Applying super-resolution to panoramic mosaics. In *WACV '98: Proceedings of the 4th IEEE Workshop on Applications of Computer Vision (WACV'98)*, page 286. IEEE Computer Society, 1998.

[24] D. Zorin and A. Barr. Correction of geometric perceptual distortions in pictures. In *Proceedings of SIGGRAPH 1995, ACM SIGGRAPH*, pages 257–264, 1995.