Appearing in Proc. Workshop on Representations of Visual Scenes, Cambridge MA, 1995.

# Physically-Valid View Synthesis by Image Interpolation

Steven M. Seitz
seitz@cs.wisc.edu

Charles R. Dyer
dyer@cs.wisc.edu

Department of Computer Sciences
University of Wisconsin
Madison, WI 53706

**Abstract**

*Image warping is a popular tool for smoothly transforming one image to another. "Morphing" techniques based on geometric image interpolation create compelling visual effects, but the validity of such transformations has not been established. In particular, does 2D interpolation of two views of the same scene produce a sequence of physically valid in-between views of that scene? In this paper, we describe a simple image rectification procedure which guarantees that interpolation does in fact produce valid views, under generic assumptions about visibility and the projection process. Towards this end, it is first shown that two basis views are sufficient to predict the appearance of the scene within a specific range of new viewpoints. Second, it is demonstrated that interpolation of the rectified basis images produces exactly this range of views. Finally, it is shown that generating this range of views is a theoretically well-posed problem, requiring neither knowledge of camera positions nor 3D scene reconstruction. A scanline algorithm for view interpolation is presented that requires only four user-provided feature correspondences to produce valid orthographic views. The quality of the resulting images is demonstrated with interpolations of real imagery.*

## 1 Introduction

Despite significant advances in 3D computer graphics, the realism of rendered images is limited by hand-coded graphical models. Existing techniques for creating 3D models are time intensive and put high demands on the artistry of the modeler. In light of these limitations, there has been growing interest in the use of 2D image warping techniques for image synthesis and animation. The advantage of working in 2D is that photographs of real scenes can be used as a basis to create very realistic effects. A good example of such an effect is *morphing* which combines an interpolating warp and a cross dissolve of two images to produce in-between images. Morphing techniques have been very successful in the entertainment industry, providing a simple mechanism capable of producing visually appealing transformations from one image to another. Despite the popularity of morphing techniques, little attention has focused on the physical validity of the resulting images.

In this paper, we investigate the feasibility of using image warping techniques for *view synthesis*. We use the term *view synthesis* to refer to the rendering of images of an observed object or scene from new viewpoints. A special case of view synthesis is *view interpolation*, which concerns the synthesis of a continuous series of views starting at one known viewpoint and ending at another. In the context of these definitions, the main result of this paper is that for a broad class of scenes and images **image interpolation is a physically valid mechanism for view interpolation**. This result provides a theoretical basis for morphing techniques and demonstrates that views can be synthesized with simple 2D image operations. In addition, we demonstrate constructively that, unlike 3D structure recovery, view interpolation is well-posed and does not suffer from the aperture problem. The result depends on an assumption of *monotonicity* which requires that corresponding scene points appear in the same order in both images.

Practical applications of view synthesis include virtual teleconferencing [1, 2] with limited network bandwidth. By using view synthesis at the receiving end, different views of the participants can be synthesized from a small number of transmitted views. View synthesis has also been used to create panoramic mosaic images [3]. Several images of a scene can be combined to create a single mosaic image by warping the images to be consistent with a common viewpoint. An advantage of image-based view synthesis is that rendering time is independent of scene complexity. This property can be exploited to speed up rendering of complex scenes [4].

The remainder of the paper is structured as follows: Section 2 reviews related work in image-based view synthesis. Section 3 describes the projection model and relevant terminology. Section 4 formalizes the notion of view interpolation and proves that the problem is well-posed under a general visibility assumption. The feasibility of using image interpolation for view in-

The support of the National Science Foundation under Grant Nos. IRI-9220782 and CDA-9222948 is gratefully acknowledged.

terpolation is explored in Section 5 and a scanline algorithm for view interpolation using minimal correspondence information is introduced in Section 6. Section 7 presents results on real images.

## 2 Related Work

Ullman and Basri [5] demonstrated that new views can be expressed as linear combinations of other views of the same scene. Although the focus of their work was recognition, it has clear ramifications for view synthesis, providing a simple mechanism for predicting the positions of features in new views. However, their work does not take into account visibility issues that are crucial to understanding *which* views can be synthesized.

Chen and Williams [4] described an approach for view synthesis based on linear interpolation of corresponding image points using range data to obtain correspondences. They investigated special situations in which interpolation produces valid perspective views, but concluded that the interpolated images do not in general correspond to exact perspective views.

Two groups [2, 6] have recently developed image warping techniques for perspective-correct view synthesis. Under the assumption that a complete pixel-wise correspondence is available, it is possible to predict a broad range of views. Several researchers [1, 7, 8, 9, 10] have used interpolation to produce new images without establishing the physical validity of the resulting images. In addition to computing new views, these methods can be used to interpolate images of two different objects to achieve interesting effects, although the plausibility of such transformations is difficult to assess.

The applicability of each of these previous approaches is limited by the requirement that complete correspondence information must be available. A complete correspondence is generally impossible to obtain automatically, due to the aperture problem. A theoretical contribution of this paper is to show that for a general class of scenes and views, the problem of view synthesis is in fact well-posed and does not require a full correspondence.

## 3 Viewing Geometry

Under an orthographic projection model (e.g., weak perspective, paraperspective, affine), a view represents a plane onto which the scene projects to produce an image. Therefore, a view $V$ can be specified as a tuple $V = \langle \mathbf{X}, \mathbf{Y}, \mathbf{o} \rangle$ where the 3D vectors $\mathbf{X}$ and $\mathbf{Y}$ represent the coordinate axis of the image plane and the 2D vector $\mathbf{o}$ specifies the offset of the image origin from the projected world origin. The view projection matrix is denoted

$$\mathbf{\Pi} = \left[ \begin{array}{c|c} \mathbf{X}^T & \\ \mathbf{Y}^T & \mathbf{o} \end{array} \right]$$

and the projection $\mathbf{p} = (x, y)$ of a homogeneous scene point $\mathbf{P} = (X, Y, Z, 1)$ is given by $\mathbf{p} = \mathbf{\Pi P}$. The image plane unit normal, also known as the *optical axis* or *direction of gaze* of $V$ is denoted $\mathbf{Z}$. Under strict orthographic projection, $\mathbf{X}$ and $\mathbf{Y}$ are constrained to be orthonormal, whereas in a general affine model [11] $\mathbf{X}$ and $\mathbf{Y}$ may be any two linearly independent vectors. Finally, an image is the projection of the visible scene into the view. An image can be represented as an array of pixels $I$ or a matrix of feature positions $\mathbf{I}$. If $\mathbf{S}$ is a matrix whose columns are the visible homogeneous scene points then

$$\mathbf{I} = \mathbf{\Pi S} \tag{1}$$

## 4 View Synthesis from Images

The process of rendering views of a known three dimensional scene is well-understood in the graphics community. The inverse problem, of reconstructing the scene from a collection of images, has been well-studied, but is known to be ill-posed in general due to the *aperture problem*. In this paper, we are concerned with using a set of views of a scene to synthesize new views of the same scene. Because this would seem to entail solving both problems, i.e., reconstruction and rendering, the natural conclusion would be that image-based view synthesis is also inherently ill-posed. However, this turns out not to be the case. We show in this section that image-based view synthesis is in fact a well-posed problem under a monotonic visibility constraint and is not affected by the aperture problem.

### 4.1 The Monotonicity Constraint

A fundamental difficulty in 3D reconstruction from images is the problem of establishing correspondences between points in different images. The correspondence problem is often mitigated in practice by the epipolar constraint, which states that the projection of a scene feature in one image must appear along a particular line in a second image. This constraint reduces the search for correspondences to a 1-D search along epipolar lines. Further constraints have been used to help reduce the search within epipolar lines by making assumptions about the structure of the scene. One example of such a constraint is *monotonicity* [12, 13], which requires that the relative ordering of points along epipolar lines be preserved.

Let $I_1$ and $I_2$ be two images of a scene taken from views $V_1$ and $V_2$, respectively. For brevity, quantities associated with view $V_i$ will henceforth be written with subscript $i$. Any point $\mathbf{P}$ in the scene defines an *epipolar plane* $E_{12}$ spanned by the two plane normals and passing through $\mathbf{P}$. The *epipolar lines* $l_1$ and $l_2$ are the respective intersections of $E_{12}$ with $V_1$ and $V_2$.

Monotonicity states that the projections of any two points on $E_{12}$ appear in the same order along $l_1$ and $l_2$. If this property holds for all corresponding epipolar lines in the two views then we say that monotonicity holds for $I_1$ and $I_2$. Let $\mathbf{P}$ and $\mathbf{Q}$ be two scene points on the same epipolar plane that are visible in both images. Geometrically, the constraint dictates that the line through $\mathbf{P} - \mathbf{Q}$ may not intersect the line segment $\overline{\mathbf{Z}_1\mathbf{Z}_2}$ joining the tips of the two view normals.

A useful property of monotonicity is that it extends to cover a range of views in-between $V_1$ and $V_2$. We say that a third view $V_3$ is *in-between* $V_1$ and $V_2$ if its normal $\mathbf{Z}_3$ intersects $\overline{\mathbf{Z}_1\mathbf{Z}_2}$. Because the line through $\mathbf{P}$ and $\mathbf{Q}$ intersects $\overline{\mathbf{Z}_1\mathbf{Z}_2}$ if and only if it intersects

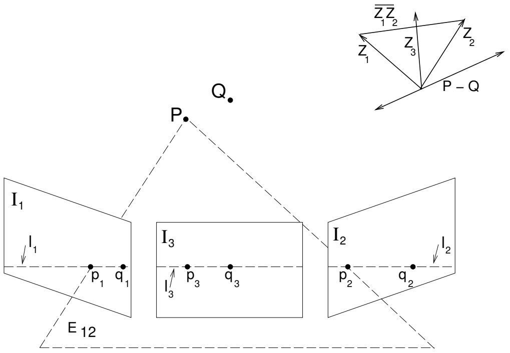


Figure 1: Monotonic Viewing Geometry. If $\mathbf{P}$ appears to the left of $\mathbf{Q}$ in images $I_1$ and $I_2$ then it must also in $I_3$, providing $\mathbf{Z}_3$ intersects $\overline{\mathbf{Z}_1\mathbf{Z}_2}$. Monotonicity requires that line $\mathbf{P}-\mathbf{Q}$ does not intersect $\overline{\mathbf{Z}_1\mathbf{Z}_2}$.

either $\overline{\mathbf{Z}_1\mathbf{Z}_3}$ or $\overline{\mathbf{Z}_3\mathbf{Z}_2}$, monotonicity of $I_1$ and $I_2$ implies monotonicity of $I_1$ and $I_3$ as well as $I_3$ and $I_2$. That is, any two points on $E_{12}$ must appear in the same order on corresponding epipolar lines of all three images. This property, that monotonicity applies to *in-between* views, is quite powerful and is sufficient to completely predict the appearance of the visible scene from all viewpoints in-between $V_1$ and $V_2$. Fig. 1 illustrates the impact of the monotonicity constraint on view synthesis.

The monotonicity condition imposes a strong visibility constraint on the scene. Intuitively, monotonicity of $I_1$ and $I_2$ means that the same scene points are visible in the range of views between $V_1$ and $V_2$. Because monotonicity is needed for view interpolation, this condition limits the set of views that can be interpolated. Nevertheless, monotonicity is satisfied at least locally for a wide range of interesting scenes.

### 4.2 The Aperture Problem

Several tasks in 3D computer vision are complicated by the *aperture problem*, which arises due to uniformly colored surfaces in the scene. In the absence of strong lighting effects, a uniform surface in the scene appears nearly uniform in projection. Although it is possible to determine which uniform regions correspond in different images, it is impossible to determine correspondences *within* these regions. As a result, additional smoothness assumptions are needed to solve problems such as optical flow and stereo vision [14]. In contrast, we show in this section that view synthesis does not suffer from the aperture problem and is therefore inherently well-posed.

Consider the projections of a set of uniform surfaces into images $I_1$ and $I_2$ (each surface is uniformly colored, but any two surfaces may have different colors). Fig. 2 depicts the cross sections $S_1$, $S_2$, and $S_3$ of three such surfaces projecting to epipolar lines $l_1$ and $l_2$. Each connected cross section projects to a uniform interval of $l_1$ and $l_2$. The monotonicity constraint induces a correspondence between the endpoints of the intervals in $l_1$ and $l_2$, determined by their relative ordering. The points on $S_1$, $S_2$, and $S_3$ projecting to the interval endpoints are determined from this correspondence by triangulation. We will refer to these scene points as *visible endpoints* of $S_1$, $S_2$, and $S_3$.

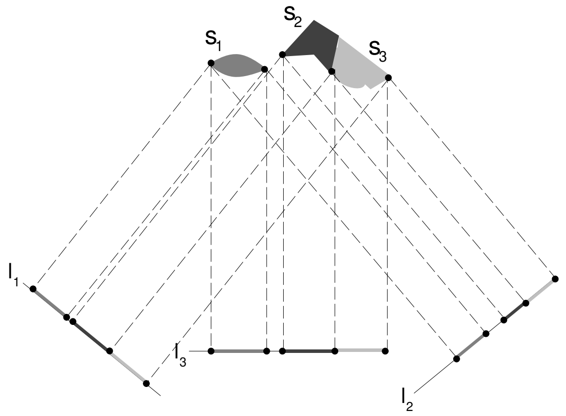


Figure 2: Correspondence Under Monotonicity. Top view of projection of three surface cross-sections into corresponding epipolar lines of images $I_1$, $I_2$, and $I_3$. Although the projected intervals in $l_1$ and $l_2$ do not provide enough information to reconstruct $S_1$, $S_2$, and $S_3$, they are sufficient to predict the appearance of $l_3$.

Now consider an in-between view $V_3$ with image $I_3$ and epipolar line $l_3$ corresponding to $l_1$ and $l_2$. $S_1$, $S_2$, and $S_3$ project to a set of uniform intervals of $l_3$, delimited by the projections of the visible endpoints of $S_1$, $S_2$, and $S_3$ . Monotonicity is needed to ensure that the endpoints of each uniform interval in $I_3$ correspond to the visible endpoints of $S_1$, $S_2$, and $S_3$.

Notice that $I_3$ does not depend on the specific shape of surfaces in the scene, only on the positions of the visible endpoints of their cross sections. Any number of distinct scenes could have produced $I_1$ and $I_2$, but each one would also project to $I_3$. Because correspondence or shape information within uniform regions is not necessary to predict in-between views, the aperture problem is avoided.

To see why the monotonicity constraint is so crucial to view synthesis, observe that it is required not only to make the correspondence problem well-posed, but also to predict the appearance of uniform surfaces whose shapes are unknown. Furthermore, the in-between views are the *only* views that can be predicted with certainty due to the requirement that the visible endpoints of each surface remain fixed. In practice, however, reasonable results may be obtained even when monotonicity is locally violated, as we demonstrate in Section 7.

## 5 View Synthesis by Image Interpolation

The previous section established that a specific range of views of a scene can be predicted from two basis views. In this section we demonstrate that knowledge of camera positions is unnecessary and that new views can be synthesized by geometrically interpolating the two basis images.

First we describe morphing techniques and discuss their application for view synthesis. Then the connections between image transformations and changes in viewpoint are discussed. It is shown that after a simple rectification procedure, linear interpolation of corresponding points in images produces new in-between views of the scene.

## 5.1 Image Interpolation

Morphing techniques combine a geometric warp with a cross dissolve to interpolate two images. A set of corresponding user-specified control points is provided in each image to guide the interpolation. As these points are typically sparse, the correspondence must be extended so that every pixel has a well-defined path. For analytical purposes, we assume for the moment that a correct and complete correspondence is provided between pixels of the two images. This constraint is relaxed in the next section to require correspondences between but not within uniform regions of the two images. We consider the common case where linear interpolation of corresponding point positions is used to create intermediate images. In other words, if $\mathbf{p}_1$ and $\mathbf{p}_2$ are corresponding points in images $I_1$ and $I_2$ respectively, the corresponding point in image $I_i$, $1 \leq i \leq 2$ is

$$\mathbf{p}_i = (2-i)\mathbf{p}_1 + (i-1)\mathbf{p}_2$$

If images $\mathbf{I}_1$ and $\mathbf{I}_2$ are represented by arrays of corresponding points, then image interpolation is expressed by the following equation:

$$\mathbf{I}_i = (2-i)\mathbf{I}_1 + (i-1)\mathbf{I}_2 \tag{2}$$

Image interpolation has a direct physical interpretation in terms of views, a connection that was recognized by Ullman and Basri [5] in the general context of linear combinations of views. Here we present a simple geometric interpretation that makes the underlying principles more explicit.

Consider two views $V_1$ and $V_2$ of a scene $\mathbf{S}$. By Eqs. (1) and (2),

$$\begin{aligned} \mathbf{I}_i &= [(2-i)\mathbf{\Pi}_1 + (i-1)\mathbf{\Pi}_2]\mathbf{S} \\ &= \mathbf{\Pi}_i \mathbf{S} \end{aligned} \tag{3}$$

where $\mathbf{\Pi}_i = (2-i)\mathbf{\Pi}_1 + (i-1)\mathbf{\Pi}_2$. $I_i$ represents what the scene would look like from a new viewpoint if every feature visible in $I_1$ and $I_2$ were also visible in $I_i$. The axes and offset of the new viewpoint are interpolations of the corresponding vectors of $V_1$ and $V_2$.

Eq. (3) provides a simple link between interpolation of images in 2D and of views in 3D. In spite of this result, image interpolations do not account for changes in visibility and often correspond to very unintuitive view interpolations. Fig. 3a graphically depicts the interpolation of views $V_1$ and $V_2$. Although both $V_1$ and $V_2$ are normal to the epipolar plane $E_{12}$, the interpolated view $V_{1.5}$ is tilted by 45 degrees with respect to $E_{12}$. In addition, the axes of $V_1$ and $V_2$ are orthonormal, whereas the axes of $V_{1.5}$ are neither orthogonal nor of unit length. Clearly, $V_{1.5}$ does not correspond to an *in-between* view, as defined in Section 4.1, so monotonicity may not be preserved and correctness of the interpolated image cannot be ensured. Furthermore, there are cases where interpolation degenerates, such as when $I_2$ is a 180 degree rotation of $I_1$. In this case, the morph collapses to a point, with all points mapping to the origin in $I_{1.5}$. In short, image interpolation will generally *not* produce valid views. Fortunately, however, these problems can be corrected by appropriately aligning the two images before performing the interpolation (see Fig. 3b), as demonstrated in the remainder of this section.

## 5.2 Image Rectification

The odd view trajectories obtained from image interpolations arise because linear interpolation of views does not amount to linear interpolation of gaze directions. Two views $V_1$ and $V_2$ each define a direction of gaze, $\mathbf{Z}_1$ and $\mathbf{Z}_2$. Intuitively, we might expect the direction of gaze to follow the most direct path between $\mathbf{Z}_1$ and $\mathbf{Z}_2$ during a smooth transition between $V_1$ and $V_2$ . However, this is generally not the case in view interpolation, as Fig. 3a illustrates, due to the nonlinear relationship between plane and normal transformations[1].

A morph can be made to interpolate gaze direction and to generate valid in-between views by first aligning the coordinate axes of the two views. This is accomplished by means of a simple image rectification procedure that aligns epipolar lines in the two images. The result of rectification is that corresponding points in the two rectified images will appear on the same scanline. In other words, a point $(x_1, y)$ in the first image will correspond to point $(x_2, y)$ in the second. The technique is a variant of the rectification procedure described in [15].

We assume that a set of at least four reference image features is provided and that their positions in each image are known. The centroid of the reference features is chosen to be the origin of each image, i.e., if $\mathbf{r}_i^1, \ldots, \mathbf{r}_i^k$ are the positions of the reference features in $I_i$ then

$$\sum_{j=1}^{k} \mathbf{r}_i^j = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Let $\mathbf{T}_i$ denote the coordinates of the top-left corner of $I_i$ in this reference frame. Denote the image coordinates of $\mathbf{r}_i^j$ as $(x_i^j, y_i^j)$. Define the measurement matrix as

$$\mathbf{M} = \begin{bmatrix} x_1^1 & \ldots & x_1^k \\ y_1^1 & \ldots & y_1^k \\ x_2^1 & \ldots & x_2^k \\ y_2^1 & \ldots & y_2^k \end{bmatrix}$$

Singular value decomposition yields the following factorization:

$$\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}$$

[1]Normals are transformed by the inverse transpose of the plane coordinate transformation.

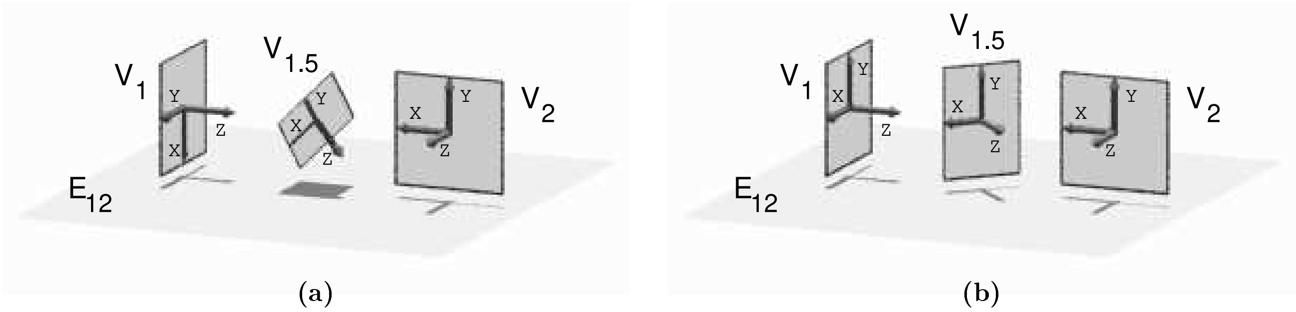


Figure 3: Views Generated by Image Interpolation. **(a)** Interpolating the $\mathbf{X}$ and $\mathbf{Y}$ axes of $V_1$ and $V_2$ produces a view that is skewed and tilted with respect to the epipolar plane $E_{12}$. **(b)** Rectification remedies the problem by aligning the view coordinate systems prior to interpolation.

Let $\mathbf{U}'$ be the matrix formed by the first 3 columns of $\mathbf{U}$. The (nonhomogeneous) affine projection matrices $\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2$ are the consecutive $2 \times 3$ blocks of $\mathbf{U}'$:

$$\left[ \begin{array}{c} \mathbf{\Pi}_1 \\ \mathbf{\Pi}_2 \end{array} \right] = \mathbf{U}'$$

The direction of epipolar lines in $I_1$ and $I_2$ can be determined from $\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2$ as follows: partition $\mathbf{\Pi}_i = [\mathbf{A}_i \mid \mathbf{d}_i]$ where $\mathbf{A}_i$ is $2 \times 2$ and $\mathbf{d}_i$ is $2 \times 1$. Define $\mathbf{B}_i$ as

$$\mathbf{B}_i = \left[ \begin{array}{cc} \mathbf{A}_i^{-1} & -\mathbf{A}_i^{-1}\mathbf{d}_i \\ 0 \;\; 0 & 1 \end{array} \right]$$

Let $\mathbf{\Pi}_1' = \mathbf{\Pi}_1\mathbf{B}_2$ and $\mathbf{\Pi}_2' = \mathbf{\Pi}_2\mathbf{B}_1$. Let $(x_i, y_i)^T$ be the third column of $\mathbf{\Pi}_i'$. In [15] it is shown that the epipolar lines in image $I_i$ make an angle of $\theta_i = arctan(\frac{y_i}{x_i})$ with the horizontal axis. To make the epipolar lines horizontal, each image is rotated by an angle of $-\theta_i$, using the matrix:

$$\mathbf{R}_{-\theta_i} = \left[ \begin{array}{cc} cos\theta_i & sin\theta_i \\ -sin\theta_i & cos\theta_i \end{array} \right]$$

To ensure that corresponding epipolar lines share the same numbered scanline, we must vertically scale $I_2$ with respect to $I_1$. Accordingly, let $\mathbf{B} = \left[ \begin{array}{c} \mathbf{R}_{-\theta_1}\mathbf{\Pi}_1' \\ (\mathbf{R}_{-\theta_2}\mathbf{\Pi}_2')_1 \end{array} \right]$ where $(\mathbf{R}_{-\theta_2}\mathbf{\Pi}_2')_1$ is the first row of $\mathbf{R}_{-\theta_2}\mathbf{\Pi}_2'$. If $\mathbf{B}$ is not invertible, either the two views have parallel optical axes or the reference features are co-planar. In either case, choose instead $\mathbf{B} = \left[ \begin{array}{c} \mathbf{R}_{-\theta_1}\mathbf{\Pi}_1' \\ 0 \;\; 0 \;\; 1 \end{array} \right]$. It follows that

$$\mathbf{R}_{-\theta_2}\mathbf{\Pi}_2'\mathbf{B}^{-1} = \left[ \begin{array}{ccc} 0 & 0 & 1 \\ 0 & s & 0 \end{array} \right]$$

If $s < 0$ it means that the epipolar lines in $I_2$ are horizontal but reversed with respect to $I_1$. In this case, $I_2$ should be rotated 180 degrees. To understand why the last column of $\mathbf{R}_{-\theta_2}\mathbf{\Pi}_2'\mathbf{B}^{-1}$ is of this form, note that the images have been rotated so that epipolar lines are horizontal. Therefore, the $y$ coordinate of a point in one image depends only upon the $y$ coordinate of the corresponding point in the other image. If the two original images are weak perspective projections, $s$ is the scaling factor of $I_2$ with respect to $I_1$. In particular, if orthographic images are used, $s$ is 1. The rectification process is completed by applying a scale matrix $\mathbf{H}_s = \left[ \begin{array}{cc} 1 & 0 \\ 0 & 1/s \end{array} \right]$. To summarize, two images $I_1$ and $I_2$ are rectified by the following sequence of image transformations:

$$\hat{I}_1 = \mathbf{R}_{-\theta_1}(I_1 + \mathbf{T}_1) \tag{4}$$

$$\hat{I}_2 = \mathbf{H}_s\mathbf{R}_{-\theta_2}(I_2 + \mathbf{T}_2) \tag{5}$$

### 5.3 Rectified View Interpolation

Although image interpolation is not always physically valid, interpolation of *rectified* monotonic images always produces valid in-between views of a scene. In light of Eq. (3), it suffices to show that rectified image interpolation preserves monotonicity. Let $\hat{V}_i$ be an interpolation of rectified views $\hat{V}_1$ and $\hat{V}_2$. Denote the first row of $\hat{\mathbf{\Pi}}_1$, $\hat{\mathbf{\Pi}}_2$, and $\hat{\mathbf{\Pi}}_i$ by $\boldsymbol{\pi}_1$, $\boldsymbol{\pi}_2$, and $\boldsymbol{\pi}_i$ respectively. Because epipolar lines are horizontal, i.e., parallel to the image x-axis, $\hat{\mathbf{X}}_1$ and $\hat{\mathbf{X}}_2$ are in the same epipolar plane. Therefore, the position of a scene point $\mathbf{P}$ along its epipolar line in $\hat{I}_j$ is given by $\boldsymbol{\pi}_j\mathbf{P}$. Let $\mathbf{P}$ and $\mathbf{Q}$ be two scene points in the same epipolar plane of $V_1$ and $V_2$. Suppose that monotonicity is satisfied for $\hat{I}_1$ and $\hat{I}_2$ so that $\boldsymbol{\pi}_1\mathbf{P} > \boldsymbol{\pi}_1\mathbf{Q}$ and $\boldsymbol{\pi}_2\mathbf{P} > \boldsymbol{\pi}_2\mathbf{Q}$. Then

$$\begin{aligned} \boldsymbol{\pi}_i(\mathbf{P} - \mathbf{Q}) &= [(2-i)\boldsymbol{\pi}_1 + (i-1)\boldsymbol{\pi}_2](\mathbf{P} - \mathbf{Q}) \\ &= (2-i)\boldsymbol{\pi}_1(\mathbf{P} - \mathbf{Q}) \\ &\quad + (i-1)\boldsymbol{\pi}_2(\mathbf{P} - \mathbf{Q}) \end{aligned} \tag{6}$$

Since both terms on the right of Eq. (6) are strictly positive for $1 \leq i \leq 2$, it follows that $\boldsymbol{\pi}_i\mathbf{P} > \boldsymbol{\pi}_i\mathbf{Q}$ and hence monotonicity is preserved.

This result indicates that linear interpolation of corresponding points in two rectified basis images always produces a valid view, assuming that monotonicity holds for the basis images. The significance of this result is that

1. Valid views can be produced by simple 2D image operations, without knowledge of either scene structure or camera geometry, and

2. Morphing techniques based on geometric image interpolation will produce physically-valid intermediate images if the basis images are appropriately rectified and satisfy monotonicity.

Image rectification ensures valid interpolated views and a smooth transition between $\hat{I}_1$ and $\hat{I}_2$. If, instead, the goal is to obtain a smooth transformation between the *original images* $I_1$ and $I_2$, it is necessary to also interpolate the rectification transformations, $\mathbf{R}_{-\theta_1}$, $\mathbf{R}_{-\theta_2}$, $\mathbf{H}_s$, $\mathbf{T}_1$, and $\mathbf{T}_2$. Accordingly, let $\hat{I}_1$ and $\hat{I}_2$ be two images rectified by Eqs. (4) and (5). The angle, scale and translation of image $I_i$ with respect to $\hat{I}_i$ are given by $\theta_i = (2-i)\theta_1 + (i-1)\theta_2$, $s_i = (2-i) + (i-1)s$, and $\mathbf{T}_i = (2-i)\mathbf{T}_1 + (i-1)\mathbf{T}_2$. Using Eqs. (4) and (5) as boundary conditions, $I_i$, $1 \leq i \leq 2$, produces a sequence of views that interpolates $I_1$ and $I_2$ given by

$$I_i = \mathbf{R}_{\theta_i}\mathbf{H}_{1/s_i}\hat{I}_i - \mathbf{T}_i \tag{7}$$

### 5.4 Orthographic View Interpolation

The above discussion demonstrates that interpolation of rectified images produces valid views using a general *affine* view model. In practice, affine projection may be too lenient in that arbitrary image skews are permitted. If the two basis images were produced by orthographic projection, what can be said about the interpolated images, i.e., are they also orthographic projections of the scene?

To address this question, suppose that $\hat{I}_1$ and $\hat{I}_2$ are rectified orthographic images of a scene with respective views $\hat{V}_1$ and $\hat{V}_2$, and $\hat{V}_i$ is an interpolated view. To see that the axes of the interpolated view are orthogonal, note that $\hat{\mathbf{X}}_1$ and $\hat{\mathbf{X}}_2$ both lie in the epipolar plane defined by $\hat{\mathbf{Z}}_1$ and $\hat{\mathbf{Z}}_2$. It follows by interpolation that $\hat{\mathbf{X}}_i$ also lies in the same epipolar plane. Because the view coordinate systems are assumed orthonormal, $\hat{\mathbf{Y}}_1$ and $\hat{\mathbf{Y}}_2$ both coincide with the epipolar plane unit normal. Therefore, $\hat{\mathbf{Y}}_i$ also coincides with the unit normal and the orthogonality of $\hat{\mathbf{X}}_i$ and $\hat{\mathbf{Y}}_i$ follows.

To determine the projective scale factors of an interpolated view, we must consider the norms of $\hat{\mathbf{X}}_i$ and $\hat{\mathbf{Y}}_i$. Although the unity of $\hat{\mathbf{Y}}_i$ is preserved, the norm of $\hat{\mathbf{X}}_i$ depends on the interior angle between gaze directions: $\theta = arccos(\hat{\mathbf{Z}}_1 \cdot \hat{\mathbf{Z}}_2)$. Specifically, it can be shown by a geometric argument that the norm of $\hat{\mathbf{X}}_i$ is given by

$$\|\hat{\mathbf{X}}_i\| = \sqrt{(2-i)^2 + 2(2-i)(i-1)cos\theta + (i-1)^2}$$

Therefore $\hat{I}_i$ is an orthographic view of the scene with aspect ratio $\|\hat{\mathbf{X}}_i\| : 1$. In general, $0 < \|\hat{\mathbf{X}}_i\| \leq 1$, with $\|\hat{\mathbf{X}}_i\|$ decreasing monotonically as $|\theta|$ increases. In particular, if $\hat{V}_1$ and $\hat{V}_2$ are within 45 degrees of one another then $\|\hat{\mathbf{X}}_i\|$ is strictly greater than 0.92. In this case, the greatest possible distortion, an 8% horizontal contraction, occurs when $i = 1.5$, corresponding to a view halfway between $\hat{I}_1$ and $\hat{I}_2$.

If $\theta$ is known, this distortion can be avoided altogether by scaling the rows of $\hat{I}_i$ by $1/\|\hat{\mathbf{X}}_i\|$. Although $\theta$ cannot be determined from the two basis images alone, any third view of the scene is sufficient to uniquely determine $\theta$ [16].

## 6 A Scanline View Interpolation Algorithm

In Section 4 we argued that synthesis of a range of views under monotonicity is possible. In this section we attest that view synthesis is practical and describe an algorithm for generating in-between views from two basis images and minimal user-provided correspondence information.

It is assumed that at least 4 corresponding reference features are provided in images $I_1$ and $I_2$. Based on these features, the images can be rectified using the procedure described in Section 5.2 to produce images $\hat{I}_1$ and $\hat{I}_2$. Once the images have been rectified, correspondences are found between uniform intervals in conjugate scanlines in the two images. With ideal data, the correspondence is completely determined by the monotonicity constraint. In practice, errors and noise in the imaging process complicate matters, causing monotonicity to be locally violated. Consequently, our approach is to find the optimal monotonic warp $\hat{W}$ of $\hat{I}_1$ that minimizes $|\hat{W}(\hat{I}_1) - \hat{I}_2|$. We employ a stereo correspondence algorithm adapted from [13] to find $W$ that uses both inter-scanline and intra-scanline constraints. We chose to use dynamic programming techniques because they make strong use of monotonicity and are relatively simple to implement. It should be noted, however, that our approach is not dependent on a particular stereo matching algorithm; other researchers [1, 9, 10] have had success with different stereo algorithms for view synthesis.

The complete algorithm is as follows:

1. Obtain either 4 or more feature correspondences or relative camera positions from the user

2. Rectify $I_1$ and $I_2$ to produce $\hat{I}_1$ and $\hat{I}_2$.

3. Match uniform intervals of corresponding scanlines in $\hat{I}_1$ and $\hat{I}_2$

4. For each scanline, linearly interpolate positions and intensities of corresponding intervals

5. Derectify $\hat{I}_i$ to produce $I_i$ using Eq. (7).

A disadvantage of this five-step approach is that it requires multiple image resampling operations, incurred by repeated rotations and scales. Since each resampling operation decreases image quality, it is advantageous to minimize the number of image transformations. One solution would be to perform steps 1 - 5 to obtain composite warping functions that directly map $I_1$ to $I_i$ and $I_2$ to $I_i$ respectively. Then the warp and cross-dissolve may be performed once at the end.

# 7 Experiments

We present the results of the algorithm applied to two views of a Band-Aid box and to two views of a stapler and cube scene. For each pair of images, 5-10 point correspondences were manually chosen. Fig. 4 illustrates the control flow of the algorithm for the Band-Aid images. The original images, $I_1$ and $I_2$, were first rectified using the procedure described in Section 5.2. A correspondence between uniform regions of $\hat{I}_1$ and $\hat{I}_2$ was found using a stereo matching algorithm. Then an image $\hat{I}_{1.5}$ halfway between $\hat{I}_1$ and $\hat{I}_2$ was produced by linear interpolation of corresponding regions. Finally, $\hat{I}_{1.5}$ was derectified to produce the intermediate image $I_{1.5}$. Notice that fine details such as the word "BAND-AID" are preserved in $I_{1.5}$ despite the fact that the image has undergone a warp, a cross-dissolve, and multiple rotation, scale, and resampling operations.

Fig. 5 shows two views of a stapler and cube scene. These images pose a challenge because some regions that are visible in one image are occluded in the other. For example, a metallic surface of the stapler is visible in $I_1$ but completely occluded in $I_2$. This region appears fuzzy in $I_{1.5}$ due to the cross-dissolve between the two images. We have found that local violations of the monotonicity assumption cause only local errors in corresponding regions of the interpolated images and do not corrupt the entire view interpolation procedure. This property is reflected in $I_{1.5}$ where the occlusion of a surface of the stapler affected only a limited area in the interpolated image.

# 8 Conclusion

In this paper we investigated the feasibility of generating new views of a scene from two basis views. Under an assumption of monotonicity, it was shown that the problem is theoretically well-posed. This result is significant in light of the fact that it is not possible to fully recover the structure of the scene due to the aperture problem. Furthermore, we demonstrated that a particular range of views can be generated by linear interpolation of the basis images, if the basis images are first rectified. This result provides a theoretical basis for morphing techniques based on geometric image interpolation [1, 7, 8, 9, 10] and provides a simple way of generating new views of a scene. Finally, a scanline algorithm for interpolating two basis images was described that requires only a small number of user-provided feature correspondences. The application of the method was demonstrated on real images.

# References


[1] S. Toelg and T. Poggio, "Towards an example-based image compression architecture for video conferencing," A.I. Memo No. 1494, M.I.T., Boston, MA, June 1994.

[2] L. McMillan and G. Bishop, "Head-tracked stereoscopic display using image warping," in *Proc. SPIE 2409A*, 1995.

[3] R. Kumar, P. Anandan, and K. Hanna, "Direct recovery of shape from multiple views: A parallax based approach," in *Proc. ICPR*, pp. 685–688, 1994.

[4] S. E. Chen and L. Williams, "View interpolation for image synthesis," in *Proc. SIGGRAPH 93*, pp. 279–288, 1993.

[5] S. Ullman and R. Basri, "Recognition by linear combinations of models," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 13, no. 10, pp. 992–1006, 1991.

[6] S. Laveau and O. Faugeras, "3-D scene representation as a collection of images and fundamental matrices," Tech. Rep. 2205, INRIA, Sophia-Antipolis, France, February 1994.

[7] G. Wolberg, *Digital Image Warping*. Los Alamitos, CA: IEEE Computer Society Press, 1990.

[8] T. Beier and S. Neely, "Feature-based image metamorphosis," in *Proc. SIGGRAPH 92*, pp. 35–42, 1992.

[9] T. Poggio and R. Brunelli, "A novel approach to graphics," A.I. Memo No. 1354, M.I.T., Boston, MA, February 1992.

[10] D. Beymer, A. Shashua, and T. Poggio, "Example based image analysis and synthesis," A.I. Memo No. 1431, M.I.T., Boston, MA, November 1993.

[11] J. J. Koenderink and A. J. van Doorn, "Affine structure from motion," *Opt. Soc. Am. A*, vol. 8, pp. 377–385, 1991.

[12] H. H. Baker and T. O. Binford, "Depth from edge and intensity based stereo," in *Proc. 7th International Joint Conference on Artificial Intelligence*, pp. 631–636, 1981.

[13] Y. Ohta and T. Kanade, "Stereo by intra- and inter-scanline search using dynamic programming," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 7, no. 2, pp. 139–154, 1985.

[14] T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory," *Nature*, vol. 317, pp. 314–319, 1985.

[15] S. M. Seitz and C. R. Dyer, "Complete structure from four point correspondences," in *Proc. Intl. Conf. on Computer Vision*, 1995. To appear.

[16] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: a factorization method," *Intl. Journal of Computer Vision*, vol. 9, no. 2, pp. 137–154, 1992.

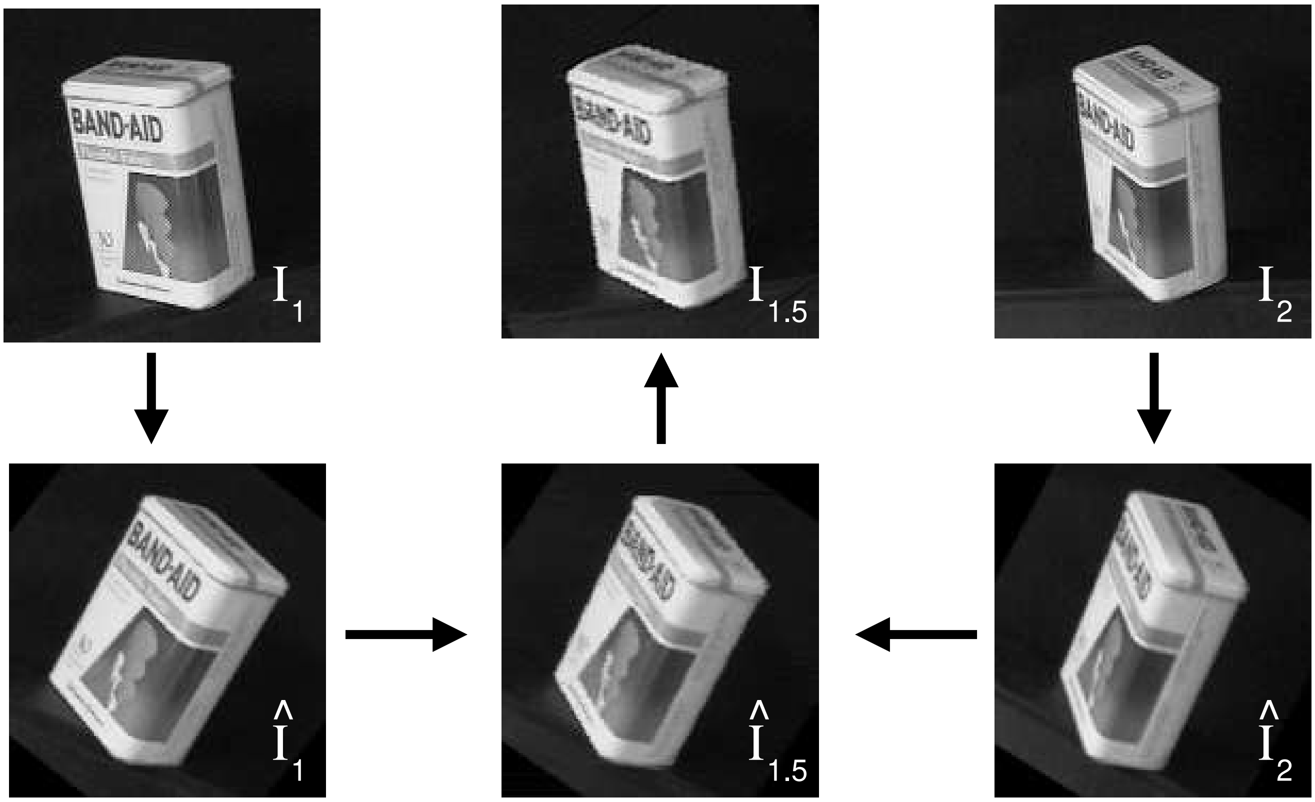


Figure 4: View interpolation control flow. The two original images are at top-left and top-right and an intermediate synthesized view is at top-center. The corresponding rectified images are shown below the originals. The arrows show the flow of the algorithm, from original to rectified to interpolated to derectified.

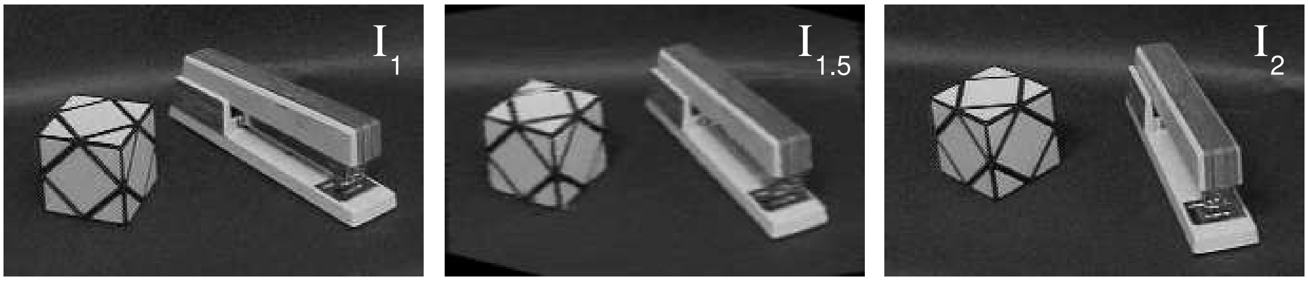


Figure 5: Interpolation of images of a cube and a stapler. Original images are at left and right, and the interpolated image is in the center. Note that a metallic surface of the stapler in the left image is occluded in the right image, locally violating the assumption of monotonicity and causing local blurring in the interpolated image. Other local artifacts, such as an incorrect region near the top of the cube in $I_{1.5}$, result from errors in correspondence.