Toward Image-Based Scene Representation Using View Morphing

Steven M. Seitz
seitz@cs.wisc.edu

Charles R. Dyer
dyer@cs.wisc.edu

Department of Computer Sciences
University of Wisconsin
Madison, WI 53706

Technical Report #1298
May 1996

Abstract

The question of which views may be inferred from a set of basis images is addressed. Under certain conditions, a discrete set of images implicitly describes scene appearance for a continuous range of viewpoints. In particular, it is demonstrated that two basis views of a static scene determine the set of all views on the line between their optical centers. Additional basis views further extend the range of predictable views to a two- or three-dimensional region of viewspace. These results are shown to apply under perspective projection subject to a generic visibility constraint called *monotonicity.* In addition, a simple scanline algorithm is presented for actually generating these views from a set of basis images. The technique, called *view morphing* may be applied to both calibrated and uncalibrated images. At a minimum, two basis views and their *fundamental matrix* are needed. Experimental results are presented on real images. This work provides a theoretical foundation for image-based representations of 3D scenes by demonstrating that perspective view synthesis is a theoretically well-posed problem.

To appear in Proc. Intl. Conf. on Pattern Recognition (ICPR 96), Vienna 1996.
The support of the National Science Foundation under Grant Nos. IRI-9220782 and CDA-9222948 is gratefully acknowledged.

1 Introduction

Image-based representations of 3D scenes are currently being developed by many researchers in the computer vision and computer graphics communities (see, for example, [12, 17, 26, 3, 11, 28]). These representations encode scene appearance with a set of images that may be adaptively combined to produce new views of a scene. Image-based techniques are especially attractive because they provide photometric information which has proven very valuable for recognition tasks [27, 18]. In addition, these representations are readily acquired from a set of basis views, avoiding the need for automatic or manual techniques for acquiring 3D object models.

At the heart of this new area lies a fundamental question: to what extent may scene appearance be modeled with a sparse set of images? Clearly, the images provide scene appearance at a discrete set of viewpoints. It is not clear, however, that a more complete coverage of viewspace is theoretically possible. A number of "view synthesis" techniques have been developed recently [12, 4, 2, 16, 29, 8, 15] to extend the range of predictable views. However, those methods require solving ill-posed correspondence tasks, suggesting that the view synthesis problem is inherently ill-posed.

As a foundation for work in this area, we feel it is necessary to answer the following two questions: given two perspective views of a static scene, under what conditions may new views be predicted? Second, *which* views are determined from a set of basis images? In this paper, we show that a specific range of perspective views is theoretically determined from two or more basis views, under a generic visibility assumption called *monotonicity*. This result applies when either the relative camera configurations are known or when only the fundamental matrix is available. In addition, we present a simple technique for generating this particular range of views using image interpolation. Importantly, the method relies only on *measurable* image information, avoiding ill-posed correspondence problems entirely. Furthermore, all processing occurs at the scanline level, effectively reducing the original 3D synthesis problem to a set of simple 1D transformations that may be implemented efficiently on existing graphics workstations. The work presented here extends to perspective projection previous results on the orthographic case [22]. In addition, this paper discusses extensions to three or more basis views, an important generalization not considered in [22].

We begin by introducing the *monotonicity* constraint and describing its implications for view synthesis in Section 2. Section 3 considers *how* views may be synthesized, and describes a simple and efficient method called *view morphing* for synthesizing new views by interpolating images, under the assumption that the relative geometry of the two cameras is known. Section 4 investigates the case where the images are *uncalibrated*, i.e., the camera geometry is unknown. Section 5 presents extensions when three or more basis views are available. Section 6 presents some results on real images.

2 View Synthesis and Monotonicity

Can the appearance from new viewpoints of a static three-dimensional scene be predicted from a set of basis views of the same scene? One way of addressing this question is to consider view synthesis as a two-step process—reconstruct the scene from the basis views using stereo or structure-from-motion methods and then reproject to form the new view. The problem with this paradigm is that view synthesis becomes at least as difficult as 3D scene reconstruction. This conclusion is especially unfortunate in light of the fact that 3D reconstruction from sparse images is generally ambiguous—a number of different scenes may be consistent with a given set of images; it is an ill-posed problem [20]. This suggests that view synthesis is also ill-posed.

In this section we present an alternate paradigm for view synthesis that avoids 3D reconstruction and dense correspondence as intermediate steps, instead relying only on *measurable* quantities, computable from a set of basis images. We first consider the conditions under which reconstruction is ill-posed and then describe why these conditions do not impede view synthesis. Ambiguity arises within regions of uniform intensity in the images. Uniform image regions provide shape and correspondence information only at boundaries. Consequently, 3D reconstruction of these regions is not possible without additional assumptions. Note however that boundary information is sufficient to predict the appearance of these regions in new views, since the region's interior is assumed to be uniform. This argument hinges on the notion that uniform regions are "preserved" in different views, a constraint formalized by the condition of *monotonicity* which we introduce next.

2.1 Notation

We write vectors and matrices in bold face and scalars in roman. Scene and image quantities are written in capitals and lowercase respectively. When possible, we also write corresponding image and scene quantities using the same letter. Images, $\mathcal{I}$, and 3D shapes or scenes, $\mathcal{S}$, are expressed as point sets. For example, an image point $(x, y) = \mathbf{p} \in \mathcal{I}$ is the projection of a scene point $(X, Y, Z) = \mathbf{P} \in \mathcal{S}$. Following convention, we represent image and scene quantities using homogeneous coordinates: a scene point with Euclidean coordinates (X, Y, Z) is expressed by the column vector $\mathbf{P} = [X\ Y\ Z\ 1]^T$ and a Euclidean image point (x, y) by $\mathbf{p} = [x\ y\ 1]^T$. We reserve the notation $\mathbf{P}$ and $\mathbf{p}$ for points expressed in Euclidean coordinates, i.e., whose last coordinate is 1. Scalar multiples of these points will be written with a tilde, as $\tilde{\mathbf{P}}$ and $\tilde{\mathbf{p}}$. A camera is represented by a 3×4 homogeneous projection matrix of the form $\mathbf{\Pi} = [\mathbf{H} \mid -\mathbf{HC}]$. The vector $\mathbf{C}$ gives the Euclidean position of the camera's optical center and the 3×3 matrix $\mathbf{H}$ specifies the position and orientation of its image plane with respect to the world coordinate system. The perspective projection equation is

$$\tilde{\mathbf{p}} = \mathbf{\Pi P} \tag{1}$$

The term *view* will henceforth refer to the tuple $\langle \mathcal{I}, \mathbf{\Pi} \rangle$ comprised of an image and its associated projection matrix.

2.2 The Monotonicity Constraint

Consider two views, V_0 and V_1, with respective optical centers $\mathbf{C}_0$ and $\mathbf{C}_1$, and images $\mathcal{I}_0$ and $\mathcal{I}_1$. Denote $\overline{\mathbf{C}_0\mathbf{C}_1}$ as the line segment connecting the two optical centers. Any point $\mathbf{P}$ in the scene determines an epipolar plane containing $\mathbf{P}$, $\mathbf{C}_0$, and $\mathbf{C}_1$ that intersects the two images in conjugate epipolar lines. The monotonicity constraint dictates that all visible scene points appear in the same order along conjugate epipolar lines of $\mathcal{I}_0$ and $\mathcal{I}_1$. This constraint is used commonly in stereo matching [1, 19] because the fixed relative ordering of points along epipolar lines simplifies the correspondence problem. Despite its usual definition with respect to epipolar lines and images, monotonicity constrains only the location of the optical centers with respect to points in the scene—the image planes may be chosen arbitrarily. An alternate definition that isolates this dependence more clearly is shown in Fig. 1. Any two scene points $\mathbf{P}$ and $\mathbf{Q}$ in the same epipolar plane determine angles θ_0 and θ_1 with the optical centers $\mathbf{C}_0$ and $\mathbf{C}_1$. The monotonicity constraint dictates that for all such points θ_0 and θ_1 must be nonzero and of equal sign. The fact that no constraint is made on the image planes is of primary importance for view synthesis because it means that *monotonicity is preserved under homographies*, i.e., under image reprojection. This fact will be essential in the next section for developing an algorithm for view synthesis.

A useful consequence of monotonicity is that it extends to cover a continuous range of views between V_0 and V_1. We say that a third view V_s is *in-between* V_0 and V_1 if its optical center $\mathbf{C}_s$ is on $\overline{\mathbf{C}_0\mathbf{C}_1}$. Observe that monotonicity is violated only when there exist two scene points, $\mathbf{P}$ and $\mathbf{Q}$, in the same

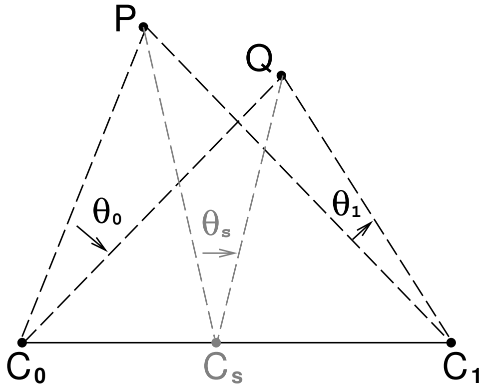

Figure 1: The Monotonicity Constraint. Any two points $\mathbf{P}$ and $\mathbf{Q}$ in the same epipolar plane determine angles θ_0 and θ_1 with the respective camera optical centers, $\mathbf{C}_0$ and $\mathbf{C}_1$. For monotonicity to apply, these angles must satisfy $\theta_0\theta_1 > 0$. If satisfied for $\mathbf{C}_0$ and $\mathbf{C}_1$, monotonicity applies as well for any other view with optical center along $\overline{\mathbf{C}_0\mathbf{C}_1}$.

epipolar plane such that the infinite line $\mathbf{PQ}$ through $\mathbf{P}$ and $\mathbf{Q}$ intersects $\overline{\mathbf{C}_0\mathbf{C}_1}$. But $\mathbf{PQ}$ intersects $\overline{\mathbf{C}_0\mathbf{C}_1}$ if and only if it intersects either $\overline{\mathbf{C}_0\mathbf{C}_s}$ or $\overline{\mathbf{C}_s\mathbf{C}_1}$. Therefore monotonicity applies to in-between views as well, i.e., signs of angles are preserved and visible scene points appear in the same order along conjugate epipolar lines of all views along $\overline{\mathbf{C}_0\mathbf{C}_1}$. We therefore refer to the range of views with centers on $\overline{\mathbf{C}_0\mathbf{C}_1}$ as a *monotonic range* of viewspace. Notice that this range gives a lower bound on the range of views for which monotonicity is satisfied in the sense that the latter set contains the former. For instance, in Fig. 1 monotonicity is satisfied for all views on the open ray from the point $\mathbf{C}_0\mathbf{C}_1 \bigcap \mathbf{PQ}$ through both camera centers. However, without *a priori* knowledge of the geometry of the scene, we may infer only that monotonicity is satisfied for the range $\overline{\mathbf{C}_0\mathbf{C}_1}$.

The 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 along $\overline{\mathbf{C}_0\mathbf{C}_1}$. Consider the projections of a set of uniform Lambertian surfaces (each surface has uniform radiance, but any two surfaces may have different radiances) into views V_0 and V_1. Fig. 2 shows cross sections $\mathcal{S}_1$, $\mathcal{S}_2$, and $\mathcal{S}_3$ of three such surfaces projecting into conjugate epipolar lines l_0 and l_1. Each connected cross section projects to a uniform interval (i.e., an interval of uniform intensity) of l_0 and l_1. The monotonicity constraint induces a correspondence between the endpoints of the intervals in l_0 and l_1, determined by their relative ordering. The points on $\mathcal{S}_1$, $\mathcal{S}_2$, and $\mathcal{S}_3$ projecting to the interval endpoints are determined from this correspondence by triangulation. We will refer to these scene points as *visible endpoints* of $\mathcal{S}_1$, $\mathcal{S}_2$, and $\mathcal{S}_3$.

Now consider an in-between view, V_s, with image $\mathcal{I}_s$ and corresponding epipolar line l_s. As a consequence of monotonicity, $\mathcal{S}_1$, $\mathcal{S}_2$, and $\mathcal{S}_3$ project to three uniform intervals along l_s, delimited by the projections of their visible endpoints. Notice that the intermediate image does not depend on the specific shapes of surfaces in the scene, only on the positions of their visible endpoints. **Any number of distinct scenes could have produced $\mathcal{I}_0$ and $\mathcal{I}_1$, but each one would also produce the same set of intermediate images.** Hence, all views along $\overline{\mathbf{C}_0\mathbf{C}_1}$ are determined from $\mathcal{I}_0$ and $\mathcal{I}_1$. This result demonstrates that view synthesis under monotonicity is an inherently well-posed problem—and is therefore much easier than 3D reconstruction and related motion analysis tasks requiring smoothness conditions and regularization techniques [20].

The monotonicity constraint is closely related to aspect graphs and visual events [9, 10]. The constraint dictates that no changes in visibility may occur within a monotonic range of viewspace. In other words,

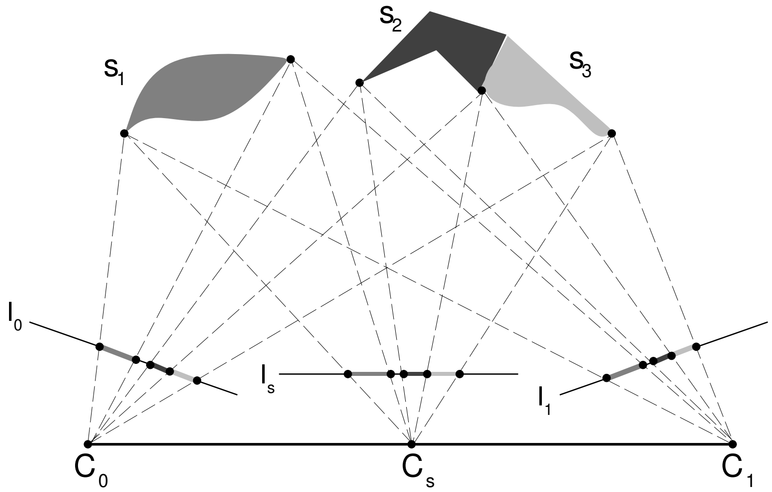

Figure 2: Correspondence Under Monotonicity. Cross-sectional view of three surfaces projecting into conjugate epipolar lines of three images. Although the projected intervals in l_0 and l_1 do not provide enough information to reconstruct $\mathcal{S}_1$, $\mathcal{S}_2$, and $\mathcal{S}_3$, they are sufficient to predict the appearance of l_s.

all views within a monotonic range are topologically equivalent; the same scene points are visible in every view. This condition of photometric topological equivalence is somewhat stronger than the notion of topological equivalence of image contour structure used to define an aspect. Consequently, a monotonic range of viewspace always occurs within an aspect.

A final question concerns the *measurability* of monotonicity. Under the monotonicity assumption we have established that view synthesis is feasible and relies only on measurable image correspondence information. However we have not yet considered whether or not monotonicity itself is measurable—can we determine if two images satisfy monotonicity by inspecting the images themselves or must we know the answer *a priori*? Strictly speaking, monotonicity is not measurable, in the sense that two images may be consistent with multiple scenes, some of which satisfy monotonicity and others that do not. However, we may determine whether or not two images are *consistent* with a scene for which monotonicity applies, by checking that each epipolar line in the first image is a monotonic warp of its conjugate in the second image. That is, if l_0 and l_1 are conjugate epipolar lines, expressed as functions mapping position to intensity, there exists a monotonic function σ such that $l_0 = l_1 \circ \sigma$. If we denote by $\mathcal{M}$ the class of all monotonic scenes consistent with two basis images, this consistency property says that we may determine from the basis images whether or not $\mathcal{M}$ is empty. If $\mathcal{M}$ is nonempty, the result of view synthesis is a set of images that are consistent with every scene in $\mathcal{M}$.

For example, consider the scene $\mathcal{S}$ shown in Fig. 2. Strictly speaking, this scene does not satisfy monotonicity for all interior points on $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{S}_3$. However, observe that $\mathcal{S}'$, the set of scene points *visible* in the two basis views, does satisfy monotonicity and therefore $\mathcal{S}' \in \mathcal{M}$. Furthermore, the visible region of $\mathcal{S}$ coincides with $\mathcal{S}'$ for all views along $\overline{\mathbf{C}_0\mathbf{C}_1}$. By the previous discussion, the appearance of $\mathcal{S}'$ and therefore $\mathcal{S}$ may be predicted for this range of views. In conclusion, even when monotonicity is not known to be satisfied, we may still synthesize views, providing the second basis image is a monotonic warp of the first.

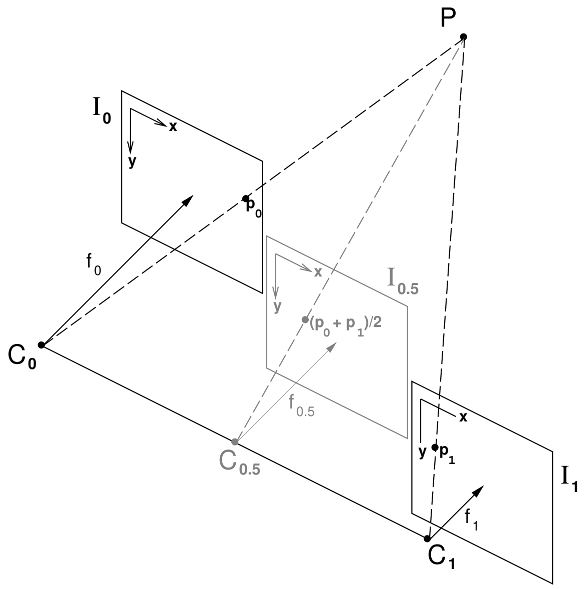

Figure 3: Morphing Parallel Views. Linear interpolation of corresponding pixels in parallel views with image planes $\mathcal{I}_0$ and $\mathcal{I}_1$ creates image $\mathcal{I}_{0.5}$, representing another parallel view of the same scene.

3 View Morphing

The previous section established that certain views are determined from two basis views under an assumption of monotonicity. In this section we present a simple approach for synthesizing these views based on image interpolation. The procedure takes as input two images, $\mathcal{I}_0$ and $\mathcal{I}_1$, their respective projection matrices, $\mathbf{\Pi}_0$ and $\mathbf{\Pi}_1$, and a third projection matrix $\mathbf{\Pi}_s$ representing the configuration of a third view along $\overline{\mathbf{C}_0\mathbf{C}_1}$. The result is a new image $\mathcal{I}_s$ representing how the visible scene appears from the third viewpoint.

3.1 Parallel Views

We begin by considering situations in which linear interpolation of images produces new views of a scene. Suppose we take a photograph $\mathcal{I}_0$ of an object, move the object in a direction parallel to the image plane of the camera, zoom out, and take a second picture $\mathcal{I}_1$, as shown in Fig. 3. Alternatively, we could produce the same two images by moving the camera instead of the object. Chen and Williams [4] previously considered this special case, arguing that linear image interpolation should produce new perspective views when the camera moves parallel to the image plane. Indeed, suppose that the camera is moved from the world origin to position $(C_X, C_Y, 0)$ and the focal length changes from f_0 to f_1. We write the respective projection matrices, $\mathbf{\Pi}_0$ and $\mathbf{\Pi}_1$, as:

$$
\begin{aligned}
\mathbf{\Pi}_0 &= \begin{bmatrix} f_0 & 0 & 0 & 0 \\ 0 & f_0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \\
\mathbf{\Pi}_1 &= \begin{bmatrix} f_1 & 0 & 0 & -f_1 C_X \\ 0 & f_1 & 0 & -f_1 C_Y \\ 0 & 0 & 1 & 0 \end{bmatrix}
\end{aligned}
$$

We refer to cameras or views with projection matrices in this form as *parallel cameras* or *parallel views*, respectively. Let $\mathbf{p}_0 \in \mathcal{I}_0$ and $\mathbf{p}_1 \in \mathcal{I}_1$ be projections of a scene point $\mathbf{P} = [X\ Y\ Z\ 1]^T$. Linear interpolation of $\mathbf{p}_0$ and $\mathbf{p}_1$ yields

$$\begin{aligned}(1-s)\mathbf{p}_0 + s\mathbf{p}_1 &= (1-s)\frac{1}{Z}\mathbf{\Pi}_0\mathbf{P} + s\frac{1}{Z}\mathbf{\Pi}_1\mathbf{P} \\ &= \frac{1}{Z}\mathbf{\Pi}_s\mathbf{P} \qquad (2)\end{aligned}$$

where

$$\mathbf{\Pi}_s = (1-s)\mathbf{\Pi}_0 + s\mathbf{\Pi}_1 \qquad (3)$$

Image interpolation therefore produces a new view whose projection matrix, $\mathbf{\Pi}_s$, is a linear interpolation of $\mathbf{\Pi}_0$ and $\mathbf{\Pi}_1$, representing a camera with center $\mathbf{C}_s$ and focal length f_s given by:

$$\begin{aligned}\mathbf{C}_s &= (sC_X, sC_Y, 0) \qquad (4) \\ f_s &= (1-s)f_0 + sf_1 \qquad (5)\end{aligned}$$

Consequently, interpolating images from parallel cameras produces the illusion of simultaneously moving the camera on the line $\overline{\mathbf{C}_0\mathbf{C}_1}$ between the two optical centers and zooming continuously.

In fact, the above derivation relies only on the equality of the third rows of $\mathbf{\Pi}_0$ and $\mathbf{\Pi}_1$. Views satisfying this more general criterion represent a broader class of parallel views for which linear image interpolation generates new views of the scene. An interesting special case is the class of orthographic projections, i.e., projections $\mathbf{\Pi}_0$ and $\mathbf{\Pi}_1$ whose last row is [0 0 0 1]. Linear interpolation of any two orthographic views of a scene therefore produces a new orthographic view of the same scene [22].

3.2 Non-Parallel Views

Using stereo rectification techniques, the problem of computing in-between views from two arbitrary perspective views can be reduced to the case treated in Section 3.1. To this end, let $\mathcal{I}_0$ and $\mathcal{I}_1$ be two perspective views with projection matrices $\mathbf{\Pi}_0 = [\mathbf{H}_0 \mid -\mathbf{H}_0\mathbf{C}_0]$ and $\mathbf{\Pi}_1 = [\mathbf{H}_1 \mid -\mathbf{H}_1\mathbf{C}_1]$. It is convenient to choose the world coordinate system so that both $\mathbf{C}_0$ and $\mathbf{C}_1$ lie on the world X-axis, i.e., $\mathbf{C}_0 = [0\ 0\ 0]^T$ and $\mathbf{C}_1 = [C_X\ 0\ 0]^T$. The two remaining axes should be chosen in a way that reduces the distortion incurred by image reprojection. A simple choice that works well in practice is to choose the Y axis in the direction of the cross product of the two image plane normals.

In between perspective views on the line $\overline{\mathbf{C}_0\mathbf{C}_1}$ may be synthesized by first applying homographies $\mathbf{H}_0^{-1}$ and $\mathbf{H}_1^{-1}$ to convert $\mathcal{I}_0$ and $\mathcal{I}_1$ to a parallel configuration. This procedure is identical to rectification techniques used in stereo vision [6]. Given a projection matrix $\mathbf{\Pi}_s = [\mathbf{H}_s \mid -\mathbf{H}_s\mathbf{C}_s]$, with $\mathbf{C}_s$ fixed by Eq. (4), the following sequence of operations, depicted in Fig. 4, produces an image $\mathcal{I}_s$ corresponding to a view with projection matrix $\mathbf{\Pi}_s$:

1. **Prewarp:** apply homographies $\mathbf{H}_0^{-1}$ to $\mathcal{I}_0$ and $\mathbf{H}_1^{-1}$ to $\mathcal{I}_1$, producing prewarped images $\hat{\mathcal{I}}_0$ and $\hat{\mathcal{I}}_1$
2. **Morph:** form $\hat{\mathcal{I}}_s$ by linearly interpolating positions and colors of corresponding points in $\hat{\mathcal{I}}_0$ and $\hat{\mathcal{I}}_1$
3. **Postwarp:** apply $\mathbf{H}_s$ to $\hat{\mathcal{I}}_s$, yielding image $\mathcal{I}_s$

Prewarping brings the image planes into alignment without changing the optical centers of the two cameras. Morphing the prewarped images moves the optical center to $\mathbf{C}_s$. Postwarping transforms the image plane of the new view to its desired position and orientation.

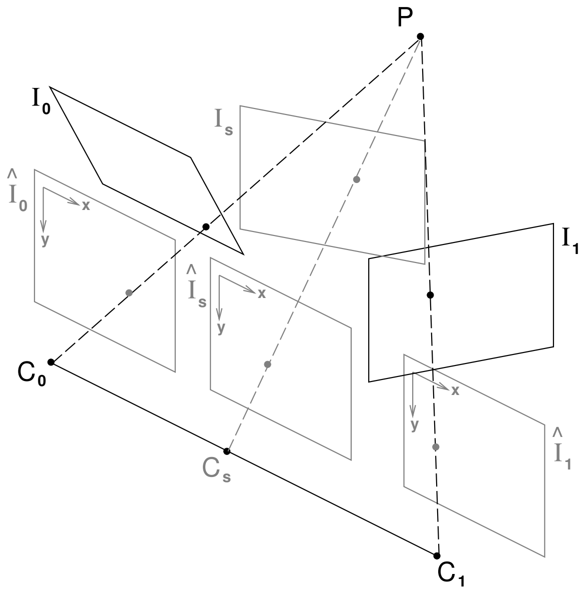

Figure 4: View Morphing in Three Steps. (1) Original images $\mathcal{I}_0$ and $\mathcal{I}_1$ are prewarped (rectified) to form parallel views $\hat{\mathcal{I}}_0$ and $\hat{\mathcal{I}}_1$. (2) $\hat{\mathcal{I}}_s$ is produced by morphing (interpolating) the prewarped images. (3) $\hat{\mathcal{I}}_s$ is postwarped to form $\mathcal{I}_s$.

Notice that the prewarped images $\hat{\mathcal{I}}_0$ and $\hat{\mathcal{I}}_1$ represent views with projection matrices $\hat{\mathbf{\Pi}}_0 = [\mathbf{I} \mid -\mathbf{C}_0]$ and $\hat{\mathbf{\Pi}}_1 = [\mathbf{I} \mid -\mathbf{C}_1]$, where $\mathbf{I}$ is the 3×3 identity matrix. Due to the special form of these projection matrices, $\hat{\mathcal{I}}_0$ and $\hat{\mathcal{I}}_1$ have the property that corresponding points in the two images appear in the same scanline. Therefore, the interpolation $\hat{\mathcal{I}}_s$ may be computed one scanline at a time using only 1D warping and resampling operations. We say that two parallel views satisfying this scanline property are in *canonical configuration*. Whereas parallelism requires that the third rows of $\mathbf{\Pi}_0$ and $\mathbf{\Pi}_1$ be equal, the scanline property requires equality of the *second* rows as well. Given projection matrices $\hat{\mathbf{\Pi}}_0 = [\mathbf{I} \mid \mathbf{0}]$ and $\hat{\mathbf{\Pi}}_1 = [\mathbf{H}_1 \mid -\mathbf{H}_1\mathbf{C}_1]$ in canonical configuration, it follows that $\mathbf{H}_1$ and $\mathbf{C}_1$ have the form:

$$\mathbf{H}_1 = \begin{bmatrix} a & b & c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{6}$$

$$\mathbf{C}_1 = [C_X \; 0 \; 0]^T \tag{7}$$

for arbitrary constants a, b, c, and C_X.

The prewarping and postwarping operations, combined with the intermediate morph, require multiple image resampling operations that may contribute to a noticeable blurring in the in-between images. Resampling effects can be reduced by supersampling the input images [30] or by composing the prewarp, morph, and postwarp transformations into one aggregate warp for each image. A disadvantage of the latter approach is that the scanline property no longer applies.

Rectification is possible providing that the epipoles are outside of the respective image borders. If this condition is not satisfied, it is still possible to apply the procedure if the prewarped images are never explicitly constructed, i.e., if aggregate warps are used. The prewarp step implicitly requires selection of a particular epipolar plane on which to reproject the basis images. Although the particular plane can be chosen arbitrarily, certain planes may be more suitable due to image sampling considerations.

Methods of choosing the rectification parameters that minimize image distortion with uniform sampling are discussed in [21].

By varying s from 0 to 1, the three-step method provides a way of generating any new view on $\overline{\mathbf{C}_0\mathbf{C}_1}$—the range of views theoretically determined by the monotonicity assumption. Notice, however, that the interpolation argument does not seem to depend on the condition that $s \in [0, 1]$, suggesting that additional views could be synthesized by choosing other values of s. Certainly, choosing any value of $s \in \Re$ will result in a new image. This image will represent a valid view of the scene only if monotonicity is preserved for $\mathbf{C}_s$. In short, for values of $s \in [0, 1]$ the resulting image is guaranteed to be valid. For values of s outside this range, the resulting image may or may not be valid, depending upon the structure of the scene.

4 Uncalibrated View Morphing

So far, we have assumed that the Euclidean camera configurations for the two basis views and the synthesized view are known. In this section we consider the case where only the basis images and the *fundamental matrix* are provided. The fundamental matrix of two views is the 3×3, rank-two matrix $\mathbf{F}$ satisfying the following relation [13, 14]:

$$\mathbf{p}_1^T \mathbf{F} \mathbf{p}_0 = 0$$

for any pair of points $\mathbf{p}_0$ and $\mathbf{p}_1$ in the two images corresponding to the same scene point. $\mathbf{F}$ is defined up to a scale factor and can be computed from the images themselves when at least 8 point correspondences are known (see [14, 7] for methods of computing $\mathbf{F}$ from point correspondences).

4.1 Rectifying Uncalibrated Images

In order to use the three-step algorithm presented in Section 3, we must find a way to rectify the images without knowing the projection matrices. Towards this end, we first consider what form $\mathbf{F}$ takes when the views are in canonical configuration and then rectify the views in such a way so that $\mathbf{F}$ achieves this form.

Let $\mathcal{I}_0$ and $\mathcal{I}_1$ be two views with projection matrices $\mathbf{\Pi}_0 = [\mathbf{I} \mid \mathbf{0}]$ and $\mathbf{\Pi}_1 = [\mathbf{H}_1 \mid -\mathbf{H}_1\mathbf{C}_1]$. Without loss of generality, we have assumed the first camera to be centered at the world origin and have set the world X and Y axes to coincide with the image coordinate axes of $\mathcal{I}_0$. The epipoles are the projections of $\mathbf{C}_1$ into $\mathcal{I}_0$ and $\mathbf{C}_0$ into $\mathcal{I}_1$:

$$\begin{aligned} \mathbf{e}_0 &= \mathbf{C}_1 && (8) \\ \mathbf{e}_1 &= -\mathbf{H}_1\mathbf{C}_1 && (9) \end{aligned}$$

Given a vector $\mathbf{p} = [x\ y\ z]^T$, we introduce the notation

$$[\mathbf{p}]_\times = \begin{bmatrix} 0 & -z & y \\ z & 0 & -x \\ -y & x & 0 \end{bmatrix}$$

Following [14], the fundamental matrix may be expressed as

$$\mathbf{F} = [\mathbf{e}_1]_\times \mathbf{H}_1$$

If the two views are in canonical configuration, we may assume without loss of generality that $\mathbf{H}_1$ and $\mathbf{C}_1$ are given by Eqs. (6) and (7) respectively, and therefore $\mathbf{e}_1 = [e_x\ 0\ 0]^T$ for some constant e_x. Consequently, the fundamental matrix is given, up to scalar multiplication, by

$$\hat{\mathbf{F}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \tag{10}$$

Conversely, suppose $\mathbf{F}$ is given by Eq. (10) and the unknown projection matrices are $\mathbf{\Pi}_0 = [\mathbf{I} \mid \mathbf{0}]$ and $\mathbf{\Pi}_1 = [\mathbf{H}_1 \mid -\mathbf{H}_1\mathbf{C}_1]$. The epipoles $\mathbf{e}_0$ and $\mathbf{e}_1$ span the null spaces of $\hat{\mathbf{F}}$ and $\hat{\mathbf{F}}^T$, respectively [21]. In particular,

$$\mathbf{e}_0 = [e_x\ 0\ 0]^T$$

for some unknown constant e_x. From Eq. (8) it follows that $\mathbf{C}_1 = [e_x\ 0\ 0]^T$. For rectification, it therefore sufficies to transform $\mathcal{I}_1$ to $\mathbf{H}_1^{-1}\mathcal{I}_1$. This image transformation induces a corresponding change in the fundamental matrix, to $\mathbf{H}_1^T\hat{\mathbf{F}}$. By the preceding argument, however, the rectified fundamental matrix is fixed, up to scalar multiple, by Eq. (10). Therefore, we have the following constraint on $\mathbf{H}_1$

$$\mathbf{H}_1^T\hat{\mathbf{F}} = \hat{\mathbf{F}}$$

It follows that $\mathbf{H}_1$ has the following structure:

$$\mathbf{H}_1 = \begin{bmatrix} a & b & c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Because $\mathbf{H}_1$ and $\mathbf{C}_1$ agree with Eqs. (6) and (7), it follows that the two views are in canonical configuration.

In summary, Eq. (10) provides a necessary and sufficient condition for testing whether two views are in canonical configuration from their fundamental matrix. In order to make use of this test, we seek a pair of homographies $\mathbf{H}_0$ and $\mathbf{H}_1$ such that the rectified images $\hat{\mathcal{I}}_0 = \mathbf{H}_0^{-1}\mathcal{I}_0$ and $\hat{\mathcal{I}}_1 = \mathbf{H}_1^{-1}\mathcal{I}_1$ have the fundamental matrix given by Eq. (10). In terms of $\mathbf{F}$ the condition on $\mathbf{H}_0$ and $\mathbf{H}_1$ is

$${\mathbf{H}_1}^T\mathbf{F}\mathbf{H}_0 = \hat{\mathbf{F}} \tag{11}$$

4.2 Choosing the Homographies

There is in fact a range of homographies $\mathbf{H}_0$ and $\mathbf{H}_1$ satisfying Eq. (11), corresponding to different choices of the rectification plane. As in the calibrated case, the particular choice of homographies is not important, except in regard to sampling considerations (see Section 3.2). What follows is a simple technique that rectifies both images with two rotations. A nice property of this method is that the rectification process reduces to that in [22] when the images are orthographic. First the rectification plane is chosen indirectly by selecting its intersection with $\mathcal{I}_0$. In homogeneous coordinates, points in the image at infinity are represented by vectors of the form $[x\ y\ 0]^T$. Any such point $\mathbf{d}_0 = [d_x\ d_y\ 0]^T$ in $\mathcal{I}_0$ represents the projection of a point $\mathbf{P}$ such that the vector $\mathbf{D} = \mathbf{P} - \mathbf{C}_0$ is parallel to $\mathcal{I}_0$. Let $\mathbf{E}$ be the vector $\mathbf{C}_1 - \mathbf{C}_0$. Providing $\mathbf{D}$ and $\mathbf{E}$ are not parallel, they span a unique epipolar rectification plane. The first step towards rectification is to bring $\mathcal{I}_0$ into alignment with this plane by rotating the image about $\mathbf{D}$ so that it is parallel with $\mathbf{E}$. This step may be performed indirectly by rotating $\mathcal{I}_0$ by an angle

θ_0 about $\mathbf{d}_0$ so that the epipole, $\mathbf{e}_0 = [e_x\ e_y\ e_z]^T$, goes to infinity. In other words, we seek a rotation matrix $\mathbf{R}_{\theta_0}^{\mathbf{d}_0}$ so that $\tilde{\mathbf{e}}_0 = \mathbf{R}_{\theta_0}^{\mathbf{d}_0}\mathbf{e}_0$ has the form $\tilde{\mathbf{e}}_0 = [\tilde{e}_x\ \tilde{e}_y\ 0]^T$. The rotation matrix is given by

$$\mathbf{R}_{\theta_0}^{\mathbf{d}_0} = \begin{bmatrix} d_x^2 + (1-d_x^2)cos\ \theta_0 & d_x d_y(1-cos\ \theta_0) & d_y sin\ \theta_0 \\ d_x d_y(1-cos\ \theta_0) & d_y^2 + (1-d_y^2)cos\ \theta_0 & -d_x sin\ \theta_0 \\ -d_y sin\ \theta_0 & d_x sin\ \theta_0 & cos\ \theta_0 \end{bmatrix} \quad (12)$$

Using Eq. (12), the desired angle of rotation θ_0 is determined to be

$$\theta_0 = -\frac{\pi}{2} - tan^{-1}\left(\frac{d_y e_x - d_x e_y}{e_z}\right)$$

At this point, epipolar lines in the image $\tilde{\mathcal{I}}_0 = \mathbf{R}_{\theta_0}^{\mathbf{d}_0}\mathcal{I}_0$ are parallel. The next step is to rotate the image so that epipolar lines are horizontal. The angle of rotation and rotation matrix are:

$$\begin{aligned} \phi_0 &= -tan^{-1}\frac{\tilde{e}_y}{\tilde{e}_x} \\ \mathbf{R}_{\phi_0} &= \begin{bmatrix} cos\ \phi_0 & -sin\ \phi_0 & 0 \\ sin\ \phi_0 & cos\ \phi_0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{aligned}$$

The analogous operations are performed on $\mathcal{I}_1$; denote $\mathbf{d}_1$ as the intersection with $\mathcal{I}_1$ of the epipolar plane spanned by $\mathbf{D}$ and $\mathbf{E}$. $\mathbf{d}_1$ may be computed as follows: if $[x\ y\ z]^T = \mathbf{F}\mathbf{d}_0$ then $\mathbf{d}_1 = [-y\ x\ 0]^T$. θ_1 and ϕ_1, the rotations about $\mathbf{d}_1$ and $[0\ 0\ 1]^T$, are found as before. At this point, if we denote $\tilde{\mathbf{H}}_0 = (\mathbf{R}_{\theta_0}^{\mathbf{d}_0}\mathbf{R}_{\phi_0})^{-1}$ and $\tilde{\mathbf{H}}_1 = (\mathbf{R}_{\theta_1}^{\mathbf{d}_1}\mathbf{R}_{\phi_1})^{-1}$, the fundamental matrix is given by

$$\tilde{\mathbf{F}} = \tilde{\mathbf{H}}_1^T\mathbf{F}\tilde{\mathbf{H}}_0 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & a \\ 0 & b & c \end{bmatrix}$$

Finally, to get $\mathbf{F}$ into the form of Eq. (10), the second image is translated and vertically scaled by the matrix

$$\mathbf{T}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -a & -c \\ 0 & 0 & b \end{bmatrix}$$

In summary, two images $\mathcal{I}_0$ and $\mathcal{I}_1$ are rectified by the following sequence of transformations:

$$\begin{aligned} \hat{\mathcal{I}}_0 &= \mathbf{R}_{\phi_0}\ \mathbf{R}_{\theta_0}^{\mathbf{d}_0}\ \mathcal{I}_0 &&= \mathbf{H}_0{}^{-1}\mathcal{I}_0 \\ \hat{\mathcal{I}}_1 &= \mathbf{T}_1\ \mathbf{R}_{\phi_1}\ \mathbf{R}_{\theta_1}^{\mathbf{d}_1}\ \mathcal{I}_1 &&= \mathbf{H}_1{}^{-1}\mathcal{I}_1 \end{aligned}$$

The entire rectification process is determined by selecting $\mathbf{d}_0$. The only constraint on $\mathbf{d}_0$ is that $\mathbf{D}$ and $\mathbf{E}$ should not be parallel. Therefore, a suitable choice is to select $\mathbf{d}_0$ so that it is orthogonal to $\mathbf{e}_0$, i.e., $\mathbf{d}_0 = [-e_y\ e_x\ 0]^T$.

In fact, the rotations in depth $\mathbf{R}_{\theta_i}^{\mathbf{d}_i}$ are sufficient to make the image planes parallel. Although this is technically sufficient for prewarping, it is useful to add the additional warps to align the scanlines. This simplifies the morph step to a scanline interpolation and also avoids bottleneck problems that arise as a result of image plane rotations [30]

Note that when $\mathcal{I}_0$ and $\mathcal{I}_1$ are orthographic or weak perspective projections, the epipoles are already at infinity [25]. Therefore, θ_0 and θ_1 are both zero and the rectification procedure reduces to a 2D rotation of both images and a shear of the second. This is precisely the rectification process needed to interpolate orthographic images in [22]. As a result, the three-step algorithm reduces to the version described in [22] in the orthographic case. Other methods of rectification for uncalibrated images are explored in [21, 6].

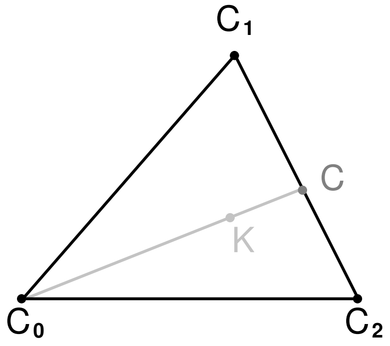

Figure 5: The View Triangle. The optical centers $\mathbf{C}_0$, $\mathbf{C}_1$, and $\mathbf{C}_3$ of three basis views determine a triangle. If monotonicity is satisfied pairwise, any view along an edge may be synthesized. If strong monotonicity applies, views in the interior may be synthesized as well. For example, a view I with optical center $\mathbf{C}$ is synthesized by interpolating the second and third basis views. An interior view at $\mathbf{K}$ is produced by a second interpolation of I and the first basis view.

4.3 Interpolating Uncalibrated Images

We have established that two images can be rectified, and therefore interpolated, without knowing their projection matrices. As in Section 3, interpolation of the prewarped images results in new views along $\overline{\mathbf{C}_0\mathbf{C}_1}$. In contrast to the calibrated case however, the postwarp step is underspecified; there is no obvious choice for the homography that transforms $\hat{I}_s$ to $\mathcal{I}_s$. One method is to simply interpolate the components of $\mathbf{H}_0^{-1}$ and $\mathbf{H}_1^{-1}$ (ϕ_i, θ_i, $\mathbf{d}_i$ and $\mathbf{T}_i$), resulting in a continuous transition from $\mathcal{I}_0$ to $\mathcal{I}_1$. We have found that a more reliable approach is to choose the postwarps so that the corners of the postwarped images linearly interpolate the corresponding corners of the original images. Whereas the first method can produce images of arbitrary size and shape, the in-between images generated by the second approach are always rectangular and interpolate the sizes of the original images. For the latter method, the postwarps are computed by determining the positions of corners points in $\hat{\mathcal{I}}_s$ and $\mathcal{I}_s$. The corners of $\mathcal{I}_s$ are found by linearly interpolating the upper-left, upper-right, lower-left, and lower-right corners of $\mathcal{I}_0$ and $\mathcal{I}_1$. The corresponding corners in $\hat{\mathcal{I}}_s$ are found by prewarping corresponding corners of $\mathcal{I}_0$ and $\mathcal{I}_1$ and linearly interpolating the resulting points. Given the positions of these four points in $\hat{\mathcal{I}}_s$ and $\mathcal{I}_s$, the required homography is found by solving a system of linear equations [30].

An alternate solution is to have the user provide the homography directly or indirectly by specification of a small number of image points [12, 24]. All three methods for choosing the postwarp transforms generally result in the synthesis of *projective* views. A projective view is a perspective view warped by a 2D affine transformation.

In conclusion, view synthesis is feasible for both calibrated and uncalibrated images using the same mechanism of rectification followed by image interpolation. When applied to calibrated images, the three-step algorithm produces new calibrated *perspective* views. If the basis images are uncalibrated, the algorithm produces *projective* views. In either case, the optical centers of the synthesized views are on the line segment between the optical centers of the basis views.

5 Three Views and Beyond

The paper up to this point has focused on image synthesis from exactly two basis views. The extension to more views is straightforward. Suppose for instance that we have three basis views that satisfy monotonicity pairwise (($\mathcal{I}_0$,$\mathcal{I}_1$), ($\mathcal{I}_0$,$\mathcal{I}_2$), and ($\mathcal{I}_1$,$\mathcal{I}_2$) each satisfy monotonicity). Three basis views permit synthesis of a triangular region of viewspace, delimited by the three optical centers. As shown in Fig. 5, each pair of basis images determines the views along one side of the triangle, spanned by $\overline{\mathbf{C}_0\mathbf{C}_1}$, $\overline{\mathbf{C}_1\mathbf{C}_2}$, and $\overline{\mathbf{C}_2\mathbf{C}_0}$.

What about interior views, i.e., views with optical centers in the interior of the triangle? Indeed, any interior view can be synthesized by a second interpolation, between a corner and a side view of the triangle. However, the assumption that monotonicity applies pairwise between corner views is not sufficient to infer monotonicity between interior views in the closed triangle $\triangle\mathbf{C}_0\mathbf{C}_1\mathbf{C}_2$; monotonicity is not transitive. In order to predict interior views, a slightly stronger constraint is needed. *Strong monotonicity* dictates that for every pair of scene points $\mathbf{P}$ and $\mathbf{Q}$, the line $\mathbf{PQ}$ does not intersect $\triangle\mathbf{C}_0\mathbf{C}_1\mathbf{C}_2$. Strong monotonicity is a direct generalization of monotonicity; in particular, strong monotonicity of $\triangle\mathbf{C}_0\mathbf{C}_1\mathbf{C}_2$ implies that monotonicity is satisfied between every pair of views centered in this triangle, and vice-versa. Consequently, strong monotonicity permits synthesis of any view in $\triangle\mathbf{C}_0\mathbf{C}_1\mathbf{C}_2$.

Now suppose we have n basis views with optical centers $\mathbf{C}_0, \ldots, \mathbf{C}_{n-1}$ and that strong monotonicity applies between each triplet of basis views[1]. By the preceding argument, any triplet of basis views determines the triangle of views between them. In particular, any view on the convex hull $\mathcal{H}$ of $\mathbf{C}_0, \ldots, \mathbf{C}_{n-1}$ is determined, as $\mathcal{H}$ is comprised of a subset of these triangles. Furthermore, the interior views are also determined: let $\mathbf{C}$ be a point in the interior of $\mathcal{H}$ and choose a corner $\mathbf{C}_i$ on $\mathcal{H}$. The line through $\mathbf{C}$ and $\mathbf{C}_i$ intersects $\mathcal{H}$ in a point $\mathbf{K}$. Since $\mathbf{K}$ lies on the convex hull, it represents the optical center of a set of views produced by two or fewer interpolations. Because $\mathbf{C}$ lies on $\overline{\mathbf{C}_i\mathbf{K}}$, all views centered at $\mathbf{C}$ are determined as well by one additional interpolation, providing monotonicity is satisfied between $\mathbf{C}_i$ and $\mathbf{K}$. To establish this last condition, observe that for monotonicity to be violated there must exist two scene points $\mathbf{P}$ and $\mathbf{Q}$ such that $\mathbf{PQ}$ intersects $\overline{\mathbf{C}_i\mathbf{K}}$, implying that $\mathbf{PQ}$ also intersects $\mathcal{H}$. Thus, $\mathbf{PQ}$ intersects at least one triangle $\triangle\mathbf{C}_i\mathbf{C}_j\mathbf{C}_k$ on $\mathcal{H}$, violating the assumption of strong monotonicity. In conclusion, n basis views determine the 3D range of viewspace contained in the convex hull of their optical centers.

This constructive argument suggests that arbitrarily large regions of viewspace may be constructed by adding more basis views. However, the prediction of any range of view-space depends on the assumption that *all* possible pairs of views within that space satisfy monotonicity. As noted in Section 2, a monotonic range may span no more than a single aspect of an aspect graph [23], thus limiting the range of views that may be predicted. Nevertheless, it is clear that a discrete set of views implicitly describes scene appearance from a continuous range of viewpoints. Based on this observation, a set of basis views is seen to constitute a scene representation, describing scene appearance as a function of viewpoint. Given an arbitrary set of basis views, the range of views that may be represented is found by partitioning the basis views into sets that obey monotonicity pairwise or strong monotonicity three at a time. Each monotonic set determines the range of views contained in its convex hull.

6 Experiments

We applied the view morphing algorithm to different pairs of basis images, two of which are shown in Fig. 6. Each pair of basis images was uncalibrated. In each case the fundamental matrix was computed from several manually-specified point correspondences. The synthesized images shown in the figure represent views halfway between the basis views.

The first pair of images represent two views of a person's face. For the most part monotonicity is satisfied, except in the region of the right ear, nose, and far sides of the face. A sparse set of user-specified

[1] In fact, strong monotonicity for each triangle on the convex hull of $\mathbf{C}_0, \ldots, \mathbf{C}_{n-1}$ is sufficient.

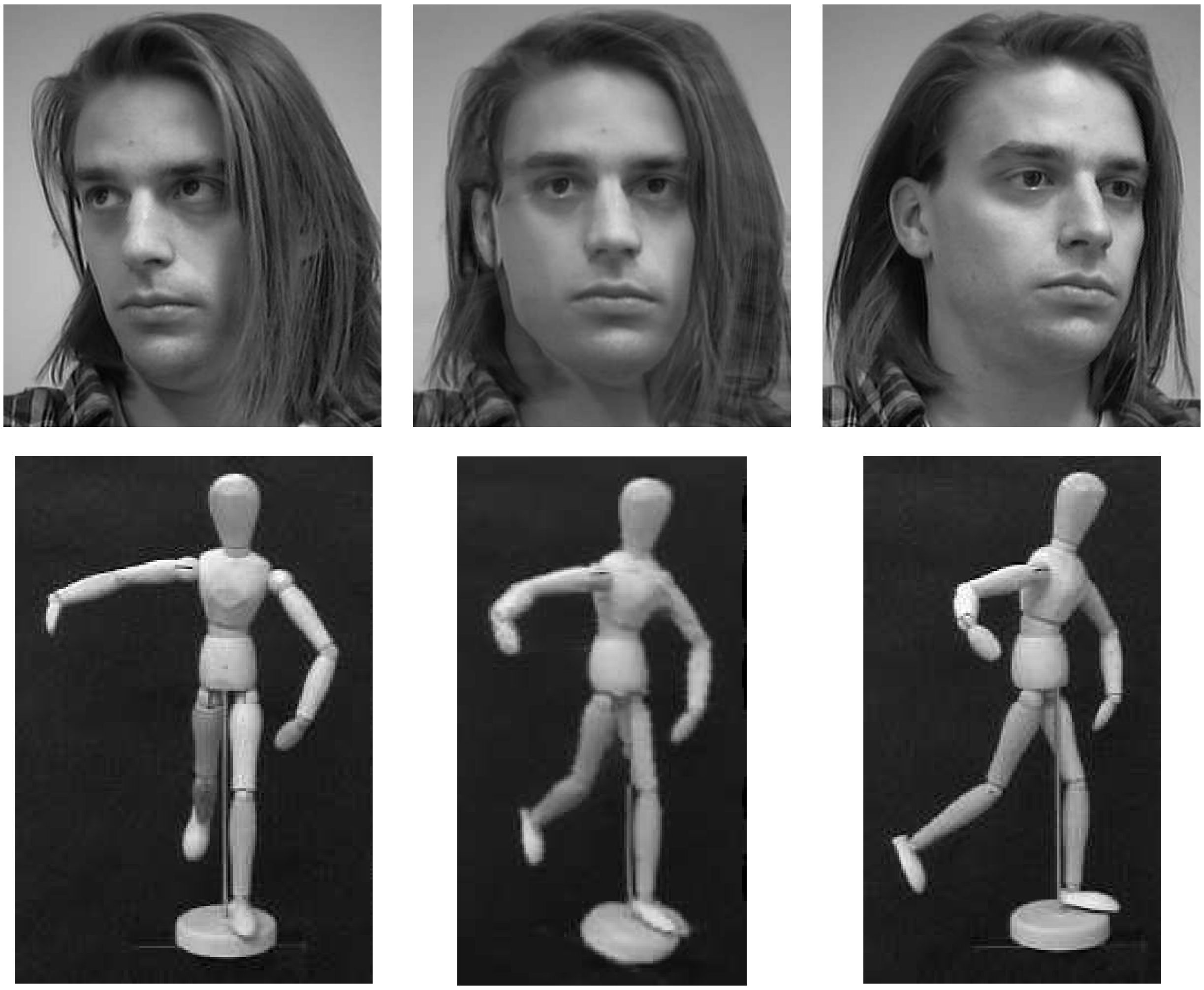

Figure 6: Morphed Views. Basis views of a face (Top) and mannequin (Bottom) are shown with halfway interpolations. The basis views appear at left and right and morphed (synthesized) images appear in the center. The morphed images use 2D image transforms to synthesize a 3D scene rotation.

feature correspondences was used to determine the correspondence map, using an image morphing technique [24]. The synthesized image represents a view from a camera viewpoint halfway between the two basis views. The image gives the convincing impression that the subject has turned his head, despite the fact that only 2D image operations have been performed. Some visible artifacts occur in regions where monotonicity has been violated, near the right ear for instance.

The second pair of images show a wooden mannequin from two viewpoints. The mannequin is an example of an object for which it is difficult to reconstruct but relatively easy to synthesize views due to lack of texture. In this example, image correspondences were automatically determined using a dynamic programming technique [19] that exploits monotonicity. Even with the monotonicity constraint, obtaining reliable correspondences with large baselines is a formidable challenge. However, incorporating limited user interaction [24] or domain knowledge [5] can significantly improve the results and is a promising line of future research.

As in the previous example, some artifacts occur where monotonicity is violated, such as near the left foot and the left thigh. Also, the synthesized view is noticeably more blurry than the basis views. Blurring is in fact evident in both synthesized views in Fig. 6, and is a direct result of image resampling. In our implementation of the view morphing algorithm, the synthesized image—a product of two projective warps and an image interpolation—is resampled three times, causing a noticeable blurring effect. The problem may be ameliorated by super-sampling the intermediate images or by concatenating the multiple image transforms into two aggregate warps and resampling only once [24].

7 Conclusion

In this paper we considered the question of which views of a static scene may be predicted from a set of two or more basis views, under perspective projection. The following results were shown:

- Under monotonicity, two perspective views determine scene appearance from the set of all viewpoints on the line between their optical centers
- Under strong monotonicity, a volume of viewspace is determined, corresponding to the convex hull of the optical centers of the basis views
- New perspective views may been synthesized by rectifying a pair of images and then interpolating corresponding pixels, one scanline at a time
- View synthesis is possible even when the views are uncalibrated, providing the *fundamental matrix* is known. In the uncalibrated case, the synthesized images represent *projective* views of the scene

These results provide a theoretical foundation for image-based representations of three-dimensional scenes, demonstrating that a discrete set of images implicitly models scene appearance for a potentially wide range of viewpoints.

References

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

[2] 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.

[3] S. E. Chen. Quicktime VR — An image-based approach to virtual environment navigation. In *Proc. SIGGRAPH 95*, pages 29–38, 1995.

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

[5] P. E. Debevec, C. J. Taylor, and J. Malik. Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach. In *Proc. SIGGRAPH 96*, 1996. To appear.

[6] O. Faugeras. *Three-Dimensional Computer Vision, A Geometric Viewpoint*. MIT Press, Cambridge, MA, 1993.

[7] R. I. Hartley. In defence of the 8-point algorithm. In *Proc. Fifth Intl. Conference on Computer Vision*, pages 1064–1070, 1995.

[8] A. Katayama, K. Tanaka, T. Oshino, and H. Tamura. A viewpoint dependent stereoscopic display using interpolation of multi-viewpoint images. In *Proc. SPIE Vol. 2409A*, pages 21–30, 1995.

[9] J. J. Koenderink and A. J. van Doorn. The singularities of the visual mapping. *Biological Cybernetics*, 24:51–59, 1976.

[10] J. J. Koenderink and A. J. van Doorn. The internal representation of solid shape with respect to vision. *Biological Cybernetics*, 32:211–216, 1979.

[11] R. Kumar, P. Anandan, M. Irani, J. Bergen, and K. Hanna. Representation of scenes from collections of images. In *Proc. IEEE Workshop on Representations of Visual Scenes*, pages 10–17, 1995.

[12] S. Laveau and O. Faugeras. 3-D scene representation as a collection of images. In *Proc. International Conference on Pattern Recognition*, pages 689–691, 1994.

[13] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.

[14] Q.-T. Luong and O. Faugeras. The fundamental matrix: Theory, algorithms, and stability analysis. *Intl. Journal of Computer Vision*, 17(1):43–75, 1996.

[15] S. Mann and R. Picard. Virtual bellows: Constructing high-quality images from video. In *Proc. First International Conference on Image Processing*, 1994.

[16] L. McMillan and G. Bishop. Head-tracked stereoscopic display using image warping. In *Proc. SPIE Vol. 2409A*, pages 21–30, 1995.

[17] L. McMillan and G. Bishop. Plenoptic modeling. In *Proc. SIGGRAPH 95*, pages 39–46, 1995.

[18] H. Murase and S. K. Nayar. Visual learning and recognition of 3-D objects from appearance. *Intl. Journal of Computer Vision*, 14(1):171–183, 1990.

[19] Y. Ohta and T. Kanade. Stereo by intra- and inter-scanline search using dynamic programming. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 7(2):139–154, 1985.

[20] T. Poggio, V. Torre, and C. Koch. Computational vision and regularization theory. *Nature*, 317:314–319, 1985.

[21] L. Robert, C. Zeller, O. Faugeras, and M. Hébert. Applications of non-metric vision to some visually guided robotics tasks. Technical Report 2584, INRIA, Sophia-Antipolis, France, June 1995.

[22] S. M. Seitz and C. R. Dyer. Physically-valid view synthesis by image interpolation. In *Proc. IEEE Workshop on Representations of Visual Scenes*, pages 18–25, 1995.

[23] S. M. Seitz and C. R. Dyer. Scene appearance representation by perspective view synthesis. Technical Report CS-TR-1298, University of Wisconsin, Madison, WI, May 1996.

[24] S. M. Seitz and C. R. Dyer. View morphing. In *Proc. SIGGRAPH 96*, 1996. To appear.

[25] L. Shapiro, A. Zisserman, and M. Brady. Motion from point matches using affine epipolar geometry. In *Proc. Third European Conference on Computer Vision*, pages 73–84, 1994.

[26] R. Szeliski. Video mosaics for virtual environments. *IEEE Computer Graphics and Applications*, 16(2):22–30, 1996.

[27] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[28] J. Y. A. Wang and E. H. Adelson. Layered representation for motion analysis. In *Proc. Computer Vision and Pattern Recognition*, pages 361–366, 1993.

[29] T. Werner, R. D. Hersch, and V. Hlavac. Rendering real-world objects using view interpolation. In *Proc. Fifth Intl. Conference on Computer Vision*, pages 957–962, 1995.

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