To appear in Proc. 13th Intl. Conf. on Pattern Recognition, Vienna, 1996.

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

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.*

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, [4, 5, 1, 3, 13]). 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. 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 [4, 2, 5, 7] 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 [10]. In addition, this paper discusses extensions to three or more basis views, an important generalization not considered in [10].

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.

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

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 [8]. 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.

Consider two views, V_0 and V_1, with respective optical centers $\mathbf{C}_0$ and $\mathbf{C}_1$, and images I_0 and 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 I_0 and I_1. This constraint is used commonly in stereo matching [6] 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.

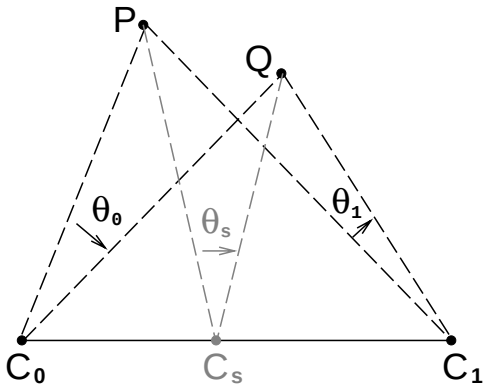


Figure 1: The Monotonicity Constraint. Any two points $\mathbf{P}$ and $\mathbf{Q}$ in the same epipolar plane determine angles θ_0 and θ_1 with the respective camera optical centers, $\mathbf{C}_0$ and $\mathbf{C}_1$. For monotonicity to apply, these angles must satisfy $\theta_0\theta_1 > 0$. If satisfied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

A useful consequence of monotonicity is that it extends to cover a continuous range of views in-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 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 S_1, S_2, and S_3 of three such surfaces projecting into conjugate epipolar lines

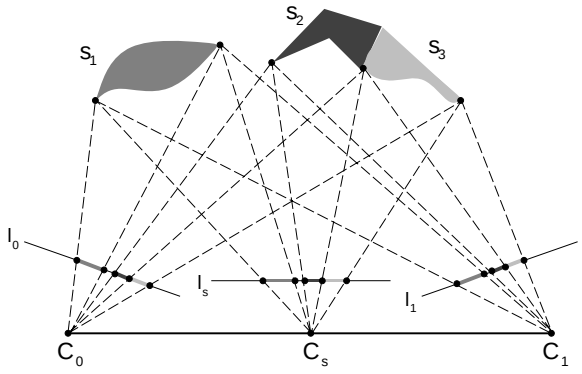


Figure 2: Correspondence Under Monotonicity. Cross-sectional view of three surfaces projecting into conjugate epipolar lines of three images. Although the projected intervals in l_0 and l_1 do not provide enough information to reconstruct S_1, S_2, and S_3, they are sufficient to predict the appearance of l_s.

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 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.

Now consider an in-between view, V_s, with image I_s and corresponding epipolar line l_s. As a consequence of monotonicity, S_1, S_2, and 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 I_0 and 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 I_0 and 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 [8].

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 $\mathcal{M}$.

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, I_0 and 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 I_s representing how the visible scene appears from the third viewpoint.

We begin with a special case where the image planes are parallel and aligned with $\overline{\mathbf{C}_0\mathbf{C}_1}$. This configuration is often used in stereo applications and will be referred to as the *parallel configuration*. The situation is expressed algebraically using the projection equations as follows. A camera is represented by a 3×4 homogeneous matrix $\mathbf{\Pi} = [\mathbf{H} \mid -\mathbf{H}\mathbf{C}]$. The optical center is given by $\mathbf{C}$ and the image plane normal is the last row of $\mathbf{H}$. A scene point (X, Y, Z) is expressed in homogeneous coordinates as $\mathbf{P} = [X\ Y\ Z\ 1]^T$ and an image point (x, y) by $\mathbf{p} = [x\ y\ 1]^T$. Because homogeneous structures are invariant under scalar multiplication, $s\mathbf{P}$ and $\mathbf{P}$ represent the same point, and similarly for $s\mathbf{p}$ and $\mathbf{p}$. We therefore reserve the notation $\mathbf{P}$ and $\mathbf{p}$ for points whose last coordinate is 1. All other multiples of these points will be denoted as $\tilde{\mathbf{P}}$ and $\tilde{\mathbf{p}}$. The perspective projection equation is:

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

In the parallel configuration, the projection matrices may be chosen so that $\mathbf{\Pi}_0 = [\mathbf{I} \mid -\mathbf{C}_0]$ and $\mathbf{\Pi}_1 = [\mathbf{I} \mid -\mathbf{C}_1]$, where $\mathbf{I}$ is the 3×3 identity matrix. Without loss of generality, we assume that $\mathbf{C}_0$ is at the world origin and $\overline{\mathbf{C}_0\mathbf{C}_1}$ is parallel to the world X-axis so that $\mathbf{C}_1 = [C_X\ 0\ 0]^T$. Let $\mathbf{p}_0$ and $\mathbf{p}_1$ be projections of a scene point $\mathbf{P} = [X\ Y\ Z\ 1]^T$ in the two views, respectively. Linear interpolation of $\mathbf{p}_0$ and $\mathbf{p}_1$ yields

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

where

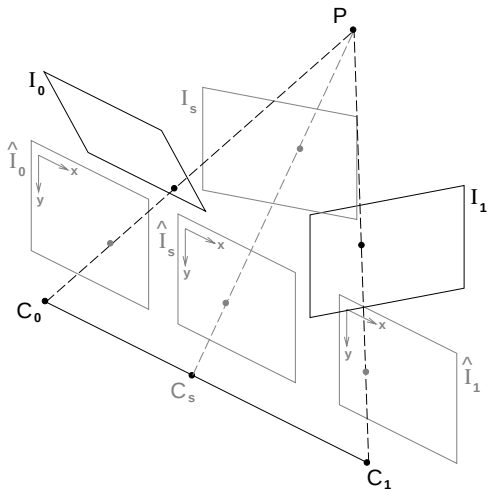


Figure 3: View Morphing in Three Steps. (1) Original images I_0 and I_1 are prewarped (rectified) to $\hat{I}_0$ and $\hat{I}_1$, in parallel configuration. (2) $\hat{I}_s$ is produced by interpolation of the prewarped images. (3) $\hat{I}_s$ is postwarped to form I_s.

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

Image interpolation, or *morphing*, therefore produces a new view whose projection matrix, $\mathbf{\Pi}_s$, is a linear interpolation of $\mathbf{\Pi}_0$ and $\mathbf{\Pi}_1$ and whose optical center is $\mathbf{C}_s = [sC_X\ 0\ 0]^T$. Eq. (1) indicates that in the parallel configuration, any parallel view along $\overline{\mathbf{C}_1\mathbf{C}_2}$ may be synthesized simply by interpolating corresponding points in the two basis views. In other words, image interpolation induces an interpolation of viewpoint for this special camera geometry.

To interpolate general views with projection matrices $\mathbf{\Pi}_0 = [\mathbf{H}_0 \mid -\mathbf{H}_0\mathbf{C}_0]$ and $\mathbf{\Pi}_1 = [\mathbf{H}_1 \mid -\mathbf{H}_1\mathbf{C}_1]$, we first apply homographies $\mathbf{H}_0^{-1}$ and $\mathbf{H}_1^{-1}$ to convert I_0 and I_1 to a parallel configuration. This procedure is identical to rectification techniques used in stereo vision [9]. This suggests a three-step procedure for view synthesis:

1. Prewarp: $\hat{I}_0 = \mathbf{H}_0^{-1} I_0$, $\hat{I}_1 = \mathbf{H}_1^{-1} I_1$
2. Morph: linearly interpolate positions and intensities of corresponding pixels in $\hat{I}_0$ and $\hat{I}_1$ to form $\hat{I}_s$
3. Postwarp: $I_s = \mathbf{H}_s \hat{I}_s$

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 the prewarp, morph, and postwarp transforms are concatenated into a pair of aggregate warps [12]. 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 [9].

4 Uncalibrated View Morphing

In order to use the view morphing algorithm presented in Section 3, we must find a way to rectify the images without knowing the projection matrices. Towards this end, it can be shown [11] that two images are in the parallel configuration when their fundamental matrix is given, up to scalar multiplication, by

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

We seek a pair of homographies $\mathbf{H}_0$ and $\mathbf{H}_1$ such that the prewarped images $\hat{I}_0 = \mathbf{H}_0^{-1} I_0$ and $\hat{I}_1 = \mathbf{H}_1^{-1} I_1$ have the fundamental matrix given by Eq. (2). In terms of $\mathbf{F}$ the condition on $\mathbf{H}_0$ and $\mathbf{H}_1$ is

$$\mathbf{H}_1{}^T \mathbf{F} \mathbf{H}_0 = \hat{\mathbf{F}} \tag{2}$$

Specific solutions to Eq. (2) are discussed in [11, 9].

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 I_s. One solution is to have the user provide the homography directly or indirectly by specification of a small number of image points [4, 12]. Another method is to simply interpolate the components of $\mathbf{H}_0^{-1}$ and $\mathbf{H}_1^{-1}$, resulting in a continuous transition from I_0 to I_1 [11]. Both 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.

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 ((I_0, I_1), (I_0, I_2), and (I_1, I_2) each satisfy monotonicity). Three basis views permit synthesis of a triangular region of viewspace, delimited by the three optical centers. 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 viewspace depends on the assumption that *all* possible pairs of views within that space satisfy monotonicity. In particular, a monotonic range may span no more than a single aspect of an aspect graph [11], 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.

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

6 Experiments

We applied the view morphing algorithm to different pairs of basis images, two of which are shown in Fig. 4. 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 feature correspondences was used to determine the correspondence map, using an image morphing technique [12]. 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 [6] that exploits monotonicity. Even with the monotonicity constraint, obtaining reliable correspondences with large baselines is a formidable challenge. However, incorporating limited user interaction [12] or domain knowledge 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. 4, 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 [12].

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. Second, under strong monotonicity, a volume of viewspace is determined, corresponding to the convex hull of the optical centers of the basis views. Third, new perspective views may

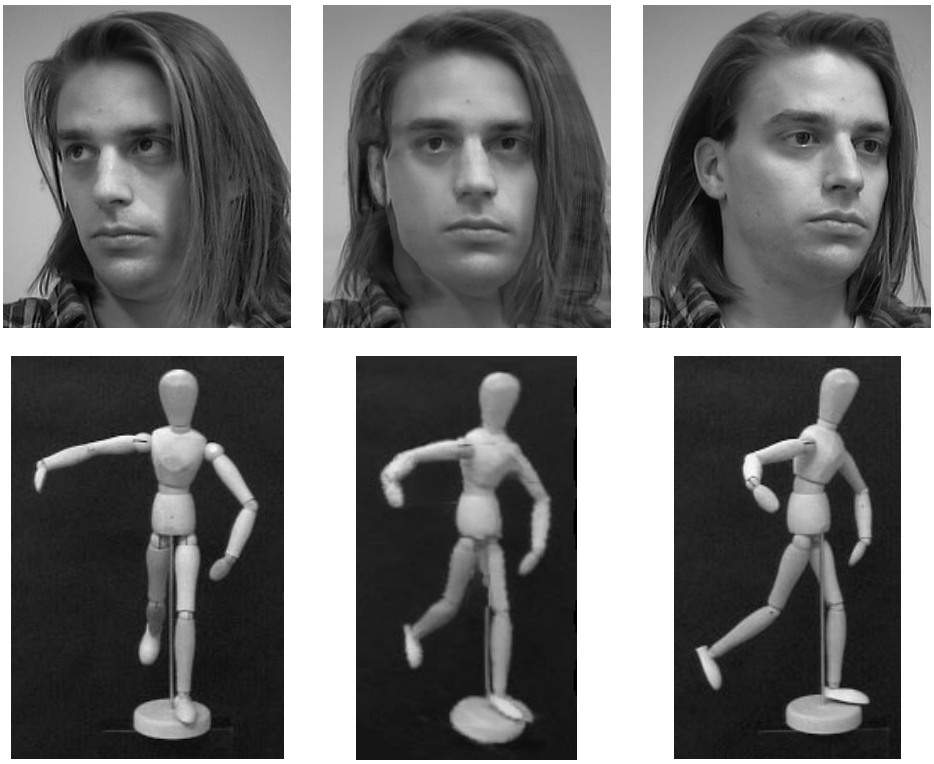


Figure 4: Morphed Views. Basis views of a face (Top) and mannequin (Bottom) are shown with halfway interpolations. The basis views appear at left and right and morphed (synthesized) images appear in the center. The morphed images use 2D image transforms to synthesize a 3D scene rotation.

be synthesized by rectifying a pair of images and then interpolating corresponding pixels, one scanline at a time, a process called *view morphing*. Fourth, 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] S. E. Chen. Quicktime VR — An image-based approach to virtual environment navigation. In *Proc. SIGGRAPH 95*, pages 29–38, 1995.

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

[3] 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.

[4] 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.

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

[6] 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.

[7] T. Poggio and D. Beymer. Learning networks for face analysis and synthesis. In *Proc. Intl. Workshop on Automatic Face- and Gesture-Recognition*, pages 160–165, Zurich, 1995.

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

[9] 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.

[10] 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.

[11] 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.

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

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