Correction of Geometric Perceptual Distortions in Pictures.

Denis Zorin, Alan H. Barr
California Institute of Technology, Pasadena, CA 91125

Abstract

We suggest an approach for correcting several types of perceived geometric distortions in computer-generated and photographic images. The approach is based on a mathematical formalization of desirable properties of pictures.

From a small set of simple assumptions we obtain perceptually preferable viewing transformations and show that these transformations can be decomposed into a perspective or parallel projection followed by a planar transformation. The decomposition is easily implemented and provides a convenient framework for further analysis of the image mapping.

We prove that two perceptually important properties are incompatible and cannot be satisfied simultaneously. It is impossible to construct a viewing transformation such that the images of all lines are straight and the images of all spheres are exact circles. Perceptually preferable tradeoffs between these two types of distortions can depend on the content of the picture. We construct parametric families of transformations with parameters representing the relative importance of the perceptual characteristics. By adjusting the settings of the parameters we can minimize the overall distortion of the picture.

It turns out that a simple family of transformations produces results that are sufficiently close to optimal. We implement the proposed transformations and apply them to computer-generated and photographic perspective projection images. Our transformations can considerably reduce distortion in wide-angle motion pictures and computer-generated animations.

Keywords: Perception, distortion, viewing transformations, perspective.

1 Introduction

The process of realistic image synthesis can be subdivided into two stages: modeling the physics of light propagation in three-dimensional environments and projecting the geometry of three-dimensional space into the picture plane (the "viewing transformation.") While the first stage is relatively independent of our understanding of visual perception, the viewing transformations are based on the fact that we are able to perceive two-dimensional patterns - pictures - as reasonably accurate representations of three-dimensional objects. We can evaluate the quality of modeling the propagation of light objectively, by comparing calculated photometric values with experimental measurements. For viewing transformations the quality is much more subjective.

a.

b.

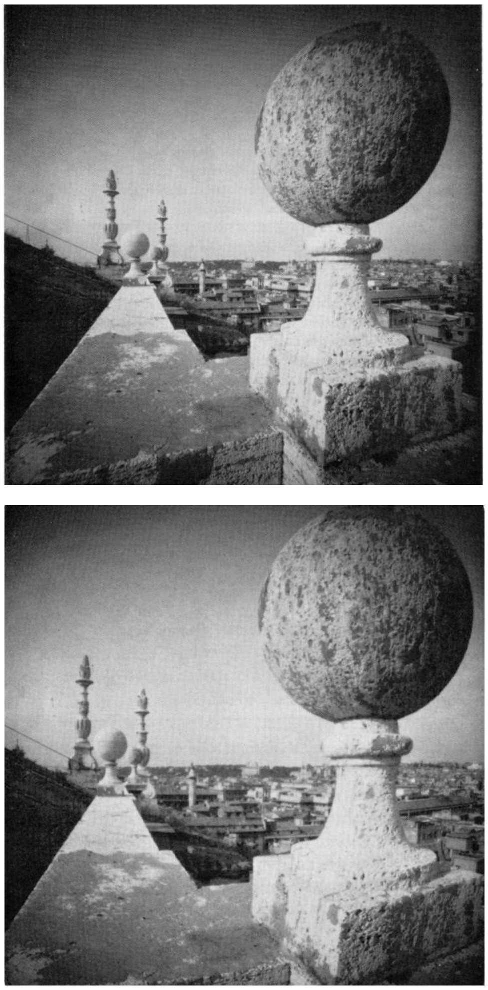

Figure 1. a. Wide-angle pinhole photograph taken on the roof of the Church of St. Ignatzio in Rome, classical example of perspective distortions from [Pir70]; reprinted with the permission of Cambridge University Press. b. Corrected version of the picture with transformation applied.

The perspective projection [1] is the viewing transformation that has been primarily used for producing realistic images, in art, photography and in computer graphics.

One motivation for using perspective projection in computer

[1] By perspective projection or *linear perspective* we mean either central projection or parallel projection into a plane.

graphics is the idea of *photorealism*: since photographic images have one of the highest degrees of realism that we can achieve, perhaps realistic rendering should model the photographic process.

But the intuitive concept of realism in many cases differs from photorealism: photographic images, are often perceived as distorted (note the shape of the sphere in Fig. 1a.) On the other hand, some paintings, while using perspective projection, contain considerable deviations from it (Fig. 2). These paintings, however, are perceptually correct and realistic.

In this paper we derive viewing transformations from some basic principles of perception of pictures rather than by modeling a particular physical process of picture generation. Our approach is based on formalizations of desirable perceptual properties of pictures as mathematical restrictions on viewing transformations.

The main result (Section 5.1) allows us to construct usable families of transformations; it is a decomposition theorem which states that under some assumptions, any perceptually acceptable picture-independent viewing transformation can be decomposed into a perspective or parallel projection and a two-dimensional transformation.

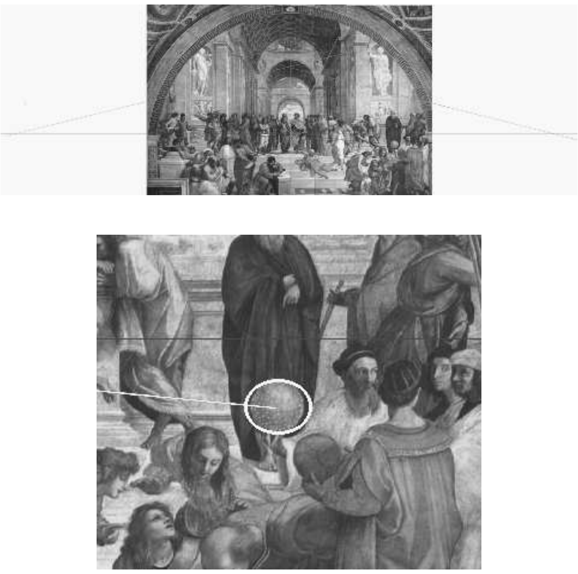

Figure 2. "School of Athens" by Rafael. (©1994-95 Christus Rex, reproduced by permission) It is possible to reconstruct the center of projection from the architectural details. The calculated image of the sphere in the right part of the picture is an ellipse with aspect ratio 1:1.2, while the painting is a perfect circle.

Our approach allows us to achieve several goals:

- We construct new viewing transformations that reduce distortions that appear in perspective projection images. It turns out that some of these transformations can be implemented as a postprocessing stage (Equation 1 , pseudocode in Section 7) after perspective projection and can be applied to existing images and photos (Figs. 1,7,8,9,10.)
- We provide a basis for understanding limitations of two-dimensional images of three-dimensional space; certain perceptual distortions can be eliminated only at the expense of increasing other distortions.
- Our transformation works well in animations and movies.
- Our families of transformations can be modified or extended by adding or removing auxiliary perceptual requirements; this provides a general basis for constructing pictures with desirable perceptual properties.

The transformations that we propose may have a number of applications: creation of computer-generated wide-angle pictures and wide-angle animations with reduced distortion, and correction of photographic images and movies.

Related Work Considerable data on picture perception have been accumulated by experimental psychologists; overviews can be found in [Kub86], [Hag80]. Computer graphics was influenced by the study of human vision in many ways: for example, RGB color representation is based on the trichromatic theory of color perception and anti-aliasing is based on various observations in visual perception.

Principles of the perception of color have been applied to computer graphics [TR93]. A curvilinear perspective system based on experimental data is described in [Rau86].

Limitations of perspective projection are well known in art and photography. [Gla94] mentions the limitations of linear perspective.

As far as we know, this paper is the first attempt to apply perceptual principles to the analysis and construction of viewing transformations in computer graphics.

Outline of the Paper. The paper is organized as follows:[2] In Section 2 we discuss the properties of linear perspective, in Section 3 we formulate our assumptions about perception of pictures and formulate some desirable properties. Section 4 describes some restrictions that we have to impose on the viewing transformation to make it practical. In Section 5.1 we discuss the decomposition theorem for viewing transformations. In Section 5.3 we discuss construction of the 2D component of decomposition,

Section 6 describes a perceptual basis for the choice of the projection component of the decomposition of viewing transformation.

In Section 7 we discuss the implementation issues and we propose some applications of our methods.

Sketches of mathematical proofs can be found in appendices in the CD-ROM version of the paper and in [Zor95].

2 Analysis of linear perspective.

The theory of linear perspective is based on the following construction (Fig. 3). Suppose the eye of an observer is located at the point O. Then, the image on the retina of his eye is created by the rays reflected from the objects in the direction of point O. If we put a plane between the observer and the scene, and paint each point on the plane with the color of the ray of light going into O and crossing the plane at this point, the image on the retina will be indistinguishable from the real scene.

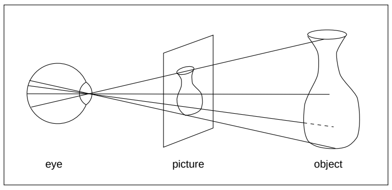

Figure 3. Pictures can produce the same retinal projection as a real object

The above argument contains some important assumptions:

- the observer looks at the scene with one eye, (or is located far enough from the image plane to consider the images in both eyes identical);
- when we look at the picture, the position of the eye coincides with the position of the eye or camera when the picture was made.

In fact, both assumptions for linear perspective are almost never true. We can look at a picture from various distances and directions

[2] The reader who is interested primarily in the implementation can go directly to Equation 1 in Section 5.3 and pseudocode in Section 7

with both eyes, but our perception of the picture doesn't change much in most cases [Hag76]. This property of pictures makes them different from illusions: while stereograms of all kinds should be observed from a particular point, traditional pictures are relatively insensitive to the changes in the viewing point. As the assumptions are not always true, it is not clear why perspective projection should be the preferred method of mapping the three-dimensional space into the plane.

In many cases we observe that perspective projection produces pictures with apparent distortions of shape and size of objects, such as distortions of shape in the margins (Figs. 1,7,8,9,10). These distortions are significantly amplified in animations and movies, resulting in shape changes of rigid bodies.

Leonardo's rule. The fact that linear perspective doesn't always produce pictures that look right was known to painters a long time ago. Leonardo da Vinci [dV70] formulated a rule which said that the best distance for depicting an object is about twenty times the size of the object. It is well-known in photography that in order to make a good portrait the camera should be placed far enough away from the object. In many paintings we can observe considerable deviations from linear perspective which in fact improve their appearance (Fig. 2.)

We conclude that there are a number of reasons to believe that linear perspective is not the only way to create realistic pictures.

3 Properties of pictures.

In this section we will describe our main assumptions about the nature of picture perception and specify the requirements that we will use in our constructions. A more detailed exposition can be found in [Zor95].

Structural features. We believe the that the features of images that are most essential for good representation are the *structural features* such as dimension (whether the image of an object is a point, a curve or an area) and presence or absence of holes or self-intersections in the image. The presence or absence of these features can be determined unambigously.

Most of the visual information that is available to the brain is contained in the images formed on the retina. We will postulate the following general requirement, which will define our concept of realistic pictures: *The retinal projections of a two-dimensional image of an object should not contain any structural features that are not present in any retinal projection of the object itself.* Structural properties of retinal images are identical to the properties of the projections into an arbitrarily chosen plane [Zor95]; our requirement can be restated in more intuitive form:

> A two-dimensional image of an object should have only structural features that are present in some planar projection of the object.

We can identify many examples of structural requirements: the image of a convex object without holes should not have holes in it, the image of a connected object cannot be disconnected, images of two intersecting objects should intersect etc. We choose a set of three structural requirements that we will use to prove the decomposition theorem in Section 5.1.

Figure 4. Mappings forbidden by structural conditions 2 and 3

1. The image of a surface should not be a point.
2. The image of a part of a straight line either shouldn't have self-intersections ("loops") or else should be a point (Fig. 4).
3. The image of a plane shouldn't have "twists" in it. This means that either each point of the plane is projected to a different point in the image, or the whole plane is projected onto a curve (Fig. 4).

We will call these conditions *structural conditions 1, 2 and 3*. Note that these requirements are quite weak: we don't require that features of some particular planar projection are represented; we just don't want to see the features that are not present in *any* projection.

Desirable properties. Next, we formulate some requirements that are not as essential as the structural ones; the corresponding features of the images can be varied continuously and can be changed within some intervals of tolerance. Examples of such features include relative sizes of objects, angles between lines, verticality. We will refer to these properties as "desirable properties." We will use two of them which we consider to be the most important. One of the most restrictive desirable properties is the following one:

Zero-curvature condition. *Images of straight lines should be straight.*

Note that this is different from the structural requirement 2 above, which is weaker. However, as we can judge straightness of lines only with some finite precision, some deviations from this property can be tolerated.

Another requirement is based on the following observation: the images of objects in the center of the picture never look distorted, given that the distance to the center of projection is large compared to the size of the object (Leonardo's rule). We will call perspective projections into the plane perpendicular to the line connecting the center of projection with the object *direct view projections*. Then the requirement eliminating distortions of shape can be stated as follows:

Direct view condition *Objects in the image should look as if they are viewed directly – as they appear in the middle of a photograph.*

Unfortunately, as we will see later, the two properties formulated above cannot be satisfied exactly at the same time.

We found several other requirements (foreshortening of objects, relative size of objects, verticality) to be of importance, but having much larger tolerance intervals. We will discuss their significance in Section 7.

4 Technical requirements

To narrow down the area of search for perceptually acceptable viewing transformations we are going to specify several additional technical requirements. They don't have any perceptual basis and are quite restrictive; however, they make the task of constructing viewing transformations manageable and the resulting transformation can be applied to a wide class of images.

1. We need a parametric family of viewing transformations so that an appropriate one can be chosen for each image.
2. The number of parameters should be small, and they must have a clear intuitive meaning.
3. The mapping must be sufficiently universal and should not depend on the details of the scene too much.

5 Derivation of viewing transformations

In the following sections we formalize the perceptual and technical conditions that were stated above and use them to prove that any viewing transformation that satisfies the structural conditions and technical conditions for any image can be implemented as a perspective projection and subsequent transformation of the picture plane. We show that direct view and zero curvature properties cannot be exactly satisfied simultaneously. We introduce quantitative measures of corresponding distortions and describe a simple parametrized family of transformations (Equation 1) where values of parameters correspond to the tradeoff between the two types of distortion. This family of transformation is close to optimal in a sense described in Section 5.3 and is easy to implement (Section 7).

5.1 General structure of viewing transformations

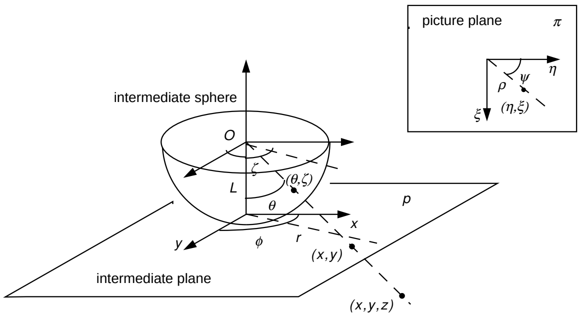

Figure 5. Coordinate systems

In this section we present a decomposition theorem derived from structural conditions 1 -3 (Section 3) and technical requirements of the previous section. The seemingly weak structural conditions 1-3 turn out to be quite restrictive if we want to construct image mappings that don't depend on the details of any particular picture, specifically, on the presence of lines or planes in any particular point of the depicted scene.

Applicaton of these requirements to *all* possible lines and planes allows us to prove that the viewing transformation with no "twists" in the images of planes and no "loops" or "folds" in the images of lines should be a composition of perspective or parallel projection and a one-to-one transformation of the plane.

In order to formulate the precise result let's introduce some definitions and notation:

We will use $x, y, ..$ for the points in the domain of a viewing transformation (a volume in 3D space), and ξ, ψ... for the points in the range (a point in the picture plane).

By a *line segment* we mean any connected subset of a line.

A viewing transformation maps many points in space to the same point in the picture plane:

Definition 1 *The set of all points of the domain of a mapping that map to a fixed point ξ is called the* **fiber** *of the mapping at the point ξ.*

In our case, fibers typically are curves in space that are mapped into single points in the picture plane.

Consider a viewing transformation which is a continuous mapping P of a region of space to the picture plane, more particularly, of an open path-connected domain $V \subset \mathbf{R}^3$ to an open subset of $\mathbf{R}^2$, satisfying the following formalizations of structural conditions:

1. The mapping of any line l to its image $P(l)$ is one-to-one everywhere or nowhere on l.
   This condition prevents "loops" in the images of lines. It is more restrictive: it doesn't allow not only "loops" but also "folds", that is, situations when each point in $P(l)$ is the image of at least two points of l.
2. The dimension of the fiber dim $P^{-1}(\xi) = 1$ for all ξ in the range of P, and all the fibers are path-connected.
   This condition prevents mapping of regions of space to single points and continuous regions to discontinuous images.
3. the mapping of a subset of a plane m to the image $P(m)$ is one-to-one everywhere or nowhere.
   This condition prevents "twists" in the images of planes.

Then our theorem can be stated as follows:

Theorem 1 *For any viewing transformation P, satisfying the conditions above, there is a perspective projection Π such that the fibers of P are subsets of fibers of Π.*

An outline of the proof is given in Appendix A.

Our practical applications are based on the following corollary:

Corollary 1 *Let a viewing transformation P satisfy the assumptions of Theorem 1 and Π be the corresponding perspective projection. If Π is a central projection, assume, in addition, that the region V lies entirely in one half-space with respect to a plane going through the center of Π. Then P can be decomposed in two ways:*

– as a composition of a perspective projection Π_{plane} into a plane followed by a transformation T_{plane} of the plane ,

– as a projection into a sphere Π_{sphere} followed by one-to-one mapping T_{sphere} of the sphere into the picture plane .

It is not true that any picture satisfying only structural conditions (without additional technical requirements) should be generated with an image mapping which has this particular decomposition, because for a particular scene the structural conditions have to be satisfied only by the objects that are actually present in it. It also should be noted that our theorem is an example of a large number of statements that can be proven given some specific choice of structural conditions. We believe that our choice is reasonable for many situations, but it is quite possible that there are cases when least restrictive requirements are sufficient and larger families of transformations can be considered.

While the choice of possible viewing transformations is considerably restricted by this theorem, there are still several degrees of freedom left:

- 2D mappings T_{plane} and T_{sphere} can be any continuous mappings.
- We can choose the center of projection for the first part of the decomposition; it is important to note that the theorem places no restrictions on the position of this center. For example, in an office scene it can be located outside the room, which is impossible for physical cameras.
- If we can split our scene into several disconnected domains (for example, foreground and background), the viewing transformation can be chosen independently for each connected part of the scene. However, separation of the space into several path connected domains introduces dependence of the transformation on the scene.

In the next sections we will consider how we can use these degrees of freedom to minimize the perceptual distortions.

5.2 Formalization of desirable conditions

In this section we formalize the conditions listed in Section 5.1, to apply them to the construction of viewing transformations. We will find error functions for both conditions that can be used as a local measure of distortion and error functionals that measure the global distortion for the whole picture.

Let's establish some notation for the viewing transformations that satisfy the conditions of the theorem.

$$\begin{tikzcd} V \subset \mathbf{R}^3 \arrow[r, "\Pi_{plane}"] \arrow[d, "\Pi_{sphere}"'] \arrow[dr, "P"] & W_{plane} \subset \mathbf{R}^2 = p \arrow[d, "T_{plane}"] \\ W_{sphere} \subset \mathbf{S}^2 \arrow[r, "T_{sphere}"] & \mathbf{R}^2 = \pi \end{tikzcd}$$

We will consider viewing transformations $P: V \subset \mathbf{R}^3 \to \mathbf{R}^2$, from an open connected domain V into the *picture plane* π, which are compositions of the projection Π_{plane} from the center O into the *intermediate plane* p and a mapping $T_{plane} : W_{plane} \subset \mathbf{R}^2 \rightarrow \mathbf{R}^2$, $W_{plane} = \Pi_{plane}(V)$, We can choose the plane p so that that the distance from O to the plane is L.

P can be also represented as a composition of central projection from O into the sphere of radius L with center at O (*intermediate sphere*) $\Pi_{sphere} : V \subset \mathbf{R}^3 \rightarrow \mathbf{S}^2$ and some mapping $T_{sphere} : W_{sphere} \subset \mathbf{S}^2 \rightarrow \mathbf{R}^2$, $W_{sphere} = \Pi_{sphere}(V)$, We will assume that the image of V in the sphere belongs to a hemisphere.

Let's introduce rectangular coordinates (x,y) and polar coordinates (r, ϕ) on the plane p; rectangular coordinates (η, ξ) and polar coordinates (ρ, ψ) in the plane π. On the sphere we will choose angular coordinate system (θ, ζ), and local coordinates in the neighborhood of a fixed point (θ_0, ζ_0): $u = L(\theta - \theta_0)$, $v = L(\zeta - \zeta_0)\sin\theta_0$ (Fig. 5).

The correspondence between T_{plane} and T_{sphere} is given by the mapping $\mathbf{S}^2 \rightarrow \mathbf{R}^2$: $\phi = \zeta$, $r = \tan\theta$.

Curvature error function Formally the restriction on images of line segments from section 3 can be expressed as follows:

Images of line segments should have zero curvature at each point.

We will call this requirement the *zero-curvature condition*. Curvature of the image of a line at a fixed point gives us a measure of how well the viewing transformation satisfies the zero-curvature condition.

If we consider the decomposition of $P = T_{plane} \circ \Pi_{plane}$, we can observe that Π_{plane} satisfies the zero-curvature condition. Therefore, we have to consider only the mapping T_{plane}. We will denote components of $T_{plane}(x, y)$, which is a point in the picture plane, by $(\eta(x, y), \xi(x, y)$

The curvature depends not only on the point but also on the direction of the line, whose image we are considering. As an error function for the zero-curvature condition at a point x we will use an estimate of the maximum curvature of the image of a line going through x.

It can be shown (see Appendix B on CD-ROM, [Zor95]) that the curvature

$$|\kappa| \leq \frac{\sqrt{|\eta_{xx}|^2 + |\eta_{yy}|^2 + 2\,|\eta_{xy}|^2 + |\xi_{xx}|^2 + |\xi_{yy}|^2 + 2\,|\xi_{xy}|^2}}{\frac{1}{2}((A + C) - \sqrt{(A - C)^2 + 4B^2})}$$

where $A = (\eta_x)^2 + (\xi_x)^2$, $B = \eta_x\eta_y + \xi_x\xi_y$, $C = (\eta_y)^2 + (\xi_y)^2$. We use a characteristic size of $T_{plane}(W)$ (corresponding to the size of the picture for perspective projection) R_0 as a scaling coefficient to obtain a dimensionless error-function. We also use the square of the curvature estimate to make the expression simpler:

$$K(T_{plane}, x, y) = R_0^2 \frac{|\eta_{xx}|^2 + |\eta_{yy}|^2 + 2\,|\eta_{xy}|^2 + |\xi_{xx}|^2 + |\xi_{yy}|^2 + 2\,|\xi_{xy}|^2}{\frac{1}{4}\left((A + C) - \sqrt{(A - C)^2 + 4B^2}\right)^2}$$

If we set $K(T_{plane}, x, y) = 0$, we can see that the all the second derivatives of η and ξ should be equal to zero, therefore, T_{plane} should be a linear transformation. This coincides with the fundamental theorem of affine geometry which says that the only transformations of the plane which map lines into lines are linear transformations.

Direct view error function. In order to formalize the direct view condition we consider mappings which are locally equivalent to direct projection as defined in Section 3. We can observe that the projection onto the sphere is locally a direct projection. Therefore, if we use the decomposition $P = T_{sphere} \circ \Pi_{sphere}$ we have to construct the mapping T_{sphere} which is locally is as close to a similarity mapping as possible. Formally, it means that the differential of the mapping T_{sphere}, which maps the tangent plane of the sphere at each point x to the plane $T_{f(x)}\mathbf{R}^2 = \mathbf{R}^2$ coinciding with the picture plane π at the point $f(x)$, should be close to a similarity mapping. The differential $DT_{f(x)}$ can be represented by the Jacobian matrix J of the mapping T_{sphere} at the point x. A nondegenerate linear transformation J is a similarity transformation if and only if $|Jw|/|w|$ doesn't depend on w.

If this ratio depends on w, then we define the direct view error function to be

$$D(T_{sphere}, \theta, \zeta) = \left| \max\left(\frac{|Jw|^2}{|w|^2}\right) / \min\left(\frac{|Jw|^2}{|w|^2}\right) - 1 \right|$$

can be used as the measure of "non-directness" of the transformation at the point (for more detailed discussion see [Zor95].).

It can be shown (see Appendix B on CD-ROM, [Zor95]) that

$$D(T_{sphere}, \theta, \zeta) = \frac{(E + G) - \sqrt{(E - G)^2 + 4F^2}}{(E + G) + \sqrt{(E - G)^2 + 4F^2}} - 1$$

where $E = (\eta_u)^2 + (\xi_u)^2$, $F = \eta_u\eta_v + \xi_u\xi_v$, $G = (\eta_v)^2 + (\xi_v)^2$.

Using the correspondence between intermediate sphere and plane we can write D as a function of T_{plane}, x and y.

Global Error functionals We want to be able to characterize the global error for each of the two perceptual requirements. We can use a norm of the error functions D and K as a measure of the global error. The choice of the norm can be different: the L^1-norm corresponds to the average error, the *sup*-norm corresponds to the maximal local error. In the first case, the error functionals are defined as

$$\mathcal{K}[T_{plane}] = \int\int_W K(T_{plane}, x, y)dxdy,$$
$$\mathcal{D}[T_{plane}] = \int\int_W D(T_{plane}, x, y)dxdy.$$

In the second case,

$$\mathcal{K}[T_{plane}] = \max_{(x,y)\in W} K(T_{plane}, x, y) \qquad \mathcal{D}[T_{plane}] = \max_{(x,y)\in W} D(T_{plane}, x, y)$$

.

5.3 2D Transformation

We are going to use the error functionals defined in Section 5.2 to construct families of transformations by solving an optimization problem. We consider only the 2D part of the decomposition of viewing transformation assuming that the perspective projection is fixed. Structural conditions suggest only that it be continuous and one-to-one.

Optimization problem. The error functionals defined above depend on the domain W_{plane} and the planar transformation T_{plane}. The first parameter is defined by the projection part of the viewing transformation, so we will assume it to be fixed now. In this case, functions satisfying $\mathcal{K}[T_{plane}] = 0$ are linear functions and the only functions satisfying $\mathcal{D}[T_{plane}] = 0$ are those derived from conformal mappings of the sphere onto the plane. Unfortunately, these two classes of functions don't intersect.

In this case for a given level μ for either error functional $\mathcal{K}$ or $\mathcal{D}$ we minimize the level of the other. The corresponding optimization problems are

$$\mathcal{K}[T_{plane}] = \min,\ \mathcal{D}[T_{plane}] = \mu \quad \text{or} \quad \mathcal{K}[T_{plane}] = \mu,\ \mathcal{D}[T_{plane}] = \min$$

These optimization problems are equivalent and can be reduced [Zor95]. to an unconstrained optimization problem for the functional $\mathcal{F}[T_{plane}] = \mu\mathcal{K}[T_{plane}] + (1 - \mu)\mathcal{D}[T_{plane}])$, where μ represents the desired tradeoff between two functionals: for $\mu = 0$ we completely ignore the zero-curvature condition, for $\mu = 1$ the direct view condition.

We also have to specify the boundary conditions in order to make the problem well-defined. This can be done by fixing the frame of the picture, that is, the values of T_{plane} on the boundary of W_{plane}.

We will consider solutions of this optimization problem in the next section.

Error functions for transformations with central symmetry From now on we will restrict our attention to transformations that also have central symmetry. This assumption allows distribution of the error evenly in all directions in the picture. The advantage of this additional restriction is a considerable simplification of the problem. The disadvantage is that real pictures seldom have this type of symmetry and, therefore, non-symmetric transformation might result in better images. We will discuss a way to create nonsymmetric transformations in Section 5.4.

In polar coordinates transformation T_{plane} can be written as $\rho = \rho(r) \quad \psi = \phi$. In this case we get the following simplified expressions for the error functions

$$K(\rho, r) = R_0^2 \frac{\frac{3}{r^2}\left(\frac{\rho}{r} - \rho'\right)^2 + \rho''^2}{\min\left(\frac{\rho^2}{r^2}, \rho'^2\right)^2}$$

$$D(\rho, r) = \frac{\max(\rho'^2(1 + r^2)^2, \rho^2(1 + \frac{1}{r^2}))}{\min(\rho'^2(1 + r^2)^2, \rho^2(1 + \frac{1}{r^2}))} - 1$$

We note that in both cases the dependence on the angular coordinate completely disappeared, so now the problem is one-dimensional. We did not use symmetry in our derivation for the general expression for $\mathcal{K}$; absence of dependence on the angle in the formulae above suggests that our bounds are quite tight.

In order for the problem to have a solution, the boundary conditions should have the same type of symmetry. We can take V to be the cone with angle at the apex θ_0. In this case $W = \Pi(V)$ will be a circle of radius $R = L \tan\theta_0$. The corresponding boundary conditions are $\rho(R) = 1$, $\rho(0) = 0$ (from continuity). Here we assume that the radius of the picture $P(V)$, corresponding to R_0 in Section 5.2, is 1.

Now there are unique functions ρ satisfying $K = 0$ or $D = 0$. For K it is obvious: $\rho_K = r/R$. For D it is $\rho_D(r) = \rho_{1D}(r)/\rho_{1D}(R)$, where $\rho_{1D}(r) = \sqrt{r^2 + 1} - 1/r$

The solutions of the optimization problem will form a parametric family $\rho(\mu, r)$ and $\rho(0, r) = \rho_D(r)$, $\rho(1, r) = \rho_K(r)$.

We consider solutions for the *sup* norm, which is more appropriate from perceptual point of view: we are guaranteed that the distortion doesn't exceed a specified amount.

Now we can state the optimization problem that we have to solve:

Minimize the functional

$$\mathcal{F}[\rho] = \max_{[0\ R]} F(\rho, r)$$

subject to boundary conditions $\rho(0) = 0$, $\rho(R) = 1$, $\rho''(0) = 0$, *where*

$$F(\rho, r) = \mu K(\rho, r) + (1 - \mu) D(\rho, r)$$

Solving a minimization problem of this type (Chebyshev minimax functional) is in general quite difficult. We found the lower estimate for the values of $\mathcal{F}$, and numerically approximated the optimal solution. It turns out that the values of $\mathcal{F}$ for linear interpolation between solutions for $\mathcal{K} = 0$ and $\mathcal{D} = 0$ are close to the optimal values.

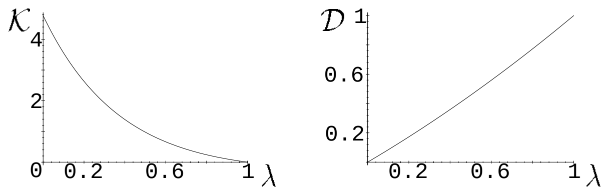

Figure 6. Functionals $\mathcal{K}$ and $\mathcal{D}$ for ρ given by Equation 1 as functions of λ for the field of view 90°. Note that when $\lambda = 1$, there is no curvature distortion. When $\lambda = 0$, then there is no distortion of shape. For a given μ, we can find λ that will approximately minimize the functional $\mathcal{F}$.

It appears that for practical purposes linear interpolation can be used. The resulting transformations have the following form:

$$\rho(r) = \lambda r/R + (1 - \lambda)\frac{R(\sqrt{r^2 + 1} - 1)}{r(\sqrt{R^2 + 1} - 1)}; \quad \psi = \phi \quad (1)$$

where the original image is represented in polar coordinate system (r, ϕ), the transformed image in polar coordinate system (ρ, ψ).

5.4 Generalization to non-symmetric cases

We can use Equation 1 to construct more general transformations by replacing a constant λ with a λ depending on the angle. In this case we can choose the balance between direct view and zero curvature conditions to be different for different directions. First, an initial constant value of λ is chosen for the whole image. Then λ is specified for a set of important directions and then interpolated for the rest of the directions. (Figure 7c). Making λ dependent on the radius and angle is more difficult, but possible; we leave this as future work.

6 Choice of viewing transformation

In the previous section we obtained an analytical expression for a family of viewing transformations parameterized by L and λ. The distance L from the center of projection to the intermediate plane p determines Π_{plane}, and λ determines the tradeoff between the zero-curvature and direct-view conditions.

We need to choose both parameters for a particular scene or image. As we have mentioned before, in our approach the center of projection need not be the position of a hypothetical camera or observer; we are free to choose it using perceptual considerations. However, we are restricted in our choice by the content of the picture that we want to obtain. In many cases, the most important constraint is the amount of *foreshortening* that we want to have across the scene. By the amount of foreshortening we mean the desired ratio of sizes of identical objects placed in the closest and most remote part of the scene (for example, human figures in the foreground and background of Fig. 9a). This ratio can be small for scenes which contain only objects of comparable size placed close to each other, such as the office scene (Fig. 9), and should be large for scenes with landscape background (Fig. 10).

According to [HEJ78] people typically prefer pictures with a small amount of foreshortening in individual objects. The behavior of the error functionals is in agreement with this fact: as we move the center of projection away from the intermediate plane ($L \to \infty$), the size of the intermediate image W_{plane} goes to 0 and it is possible to show that both direct view and zero-curvature error functionals decrease. However, a total absence of foreshortening produces distortion (Fig. 9b). The best choice of the center of projection typically corresponds to the field of view in the range 10 to 50 degrees. When such a choice is possible, we can achieve reasonably good results simply by choosing a small field of view and taking λ to be equal to 1 (Fig. 9c).

There are some types of scenes, however, that don't allow us to choose small fields of view. If we try to decrease the field of view in some scenes, either parts of the scene are lost, or the amount foreshortening becomes too close to 1 and objects in the foreground become too small. (Fig. 10c,d).

In this case, we can choose the 2D transformation by varying λ to achieve the appropriate balance between two types of distortion that we described. We choose a "global" λ for the whole image; if parts of the image still look distorted, we can make additional corrections in various parts of the image by varying λ as described in Section 5.4. (Fig 10b, Fig. 7c).

7 Implementation and Applications

Implementation. The implementation of our viewing transformations is straightforward. The Π_{plane} projection practically coincides with the standard perspective/parallel projection. There is, however, an important implementation detail that is absent in some systems. As we mentioned before, our center of projection need not coincide with the position of the camera or the eye. It is chosen according to perceptual requirements. For instance, it can happen that the most appropriate center of projection for an office scene is outside the room. In these cases it is necessary to have a mechanism for making parts of the model invisible (these parts of the model should participate in lighting calculations but should be ignored by the viewing transformation). This can be done using clipping planes.

The 2D part of the viewing transformation T_{plane} can be implemented as a separate postprocessing stage. The advantage of such an implementation is that it allows us to apply it to any perspective image, computer-generated or photographic. The only additional information required is the position of the center of projection relative to the image. The basic structure of the implementation is very simple:

```
for all output pixels (i,j) do
  r := sqrt(i^2 + j^2)
  setpixel( i,j,
      interpolated_color( ρ^{-1}(r)i/r, ρ^{-1}(r)j/r))
end
```

The inverse function ρ^{-1} can be computed numerically with any of the standard root-finding methods, such as those found in *Numerical Recipes* [PFTV88]. The interpolated_color(x, y) function computes the color for any point (x, y) with real coordinates in the original image by interpolating the colors of the integer pixels.

The position of the center of projection is usually known for computer-generated images, but is more difficult to obtain for photos. For photos it can be calculated if we know the size of the film and focal distance of the lens used in the camera. Alternatively, it can be computed directly from the image if there is a rectangular object of known aspect ratio present in the picture [Zor95].

Applications. Examples of applications of our viewing transformations were mentioned throughout the paper. We can identify the following most important applications:

- Creation of wide-angle pictures with minimal distortions. (Figs. 9,10).
- Reduction of distortions in photographic images. (Figs 7,8).
- Creation of wide-angle animation with reduced distortion of shape.
- A better alternative to fisheye views. Fisheye views are used for making images with extremely wide angle (up to 180 for hemispherical fisheye), when the distortions in linear projection make it impossible to produce any reasonable picture. However, fisheye pictures have considerable distortions of their own. The pictures that we obtain using our transformations look significantly less distorted than fisheye views.
- Zooming of parts of a wide-angle picture: For example, we can cut out a portrait of one of the authors from the transformed image in Fig. 7b, while it would look quite distorted if we had used the original photo (Fig. 7a)

8 Conclusion and Future Work

We developed an approach for constructing viewing transformations on perceptual basis. We demonstrate that two important perceptually desirable requirements are incompatible and there is no unique viewing transformation producing perceptually correct images for any scene. We described a simple family of viewing transformations suitable for reducing distortions in wide-angle images. These transformations are straightforward to implement as a postprocessing stage in a rendering system or for photographs and motion pictures.

As we have mentioned in Section 5.1, Theorem 1 applies only in cases when we consider all possible lines and planes in the scene, not only the ones present in it. Better results can be achieved by introducing direct dependence of the viewing transformation on the objects of the scene.

Possible extensions of this work include considering the dependence of λ in the Equation 1 on r, and conformal transformations of the plane preserving the direct-view property. We also can introduce new perceptually desirable properties and finding new families of transformations that produce optimal images with respect to these properties.

9 Acknowledgements

We wish to thank Bena Currin for her help with writing the software and Allen Concorran for the help with preparing the images. Many thanks to Greg Ward for his RADIANCE system that was used to render the image in the paper. We also thank the members of Caltech Computer Graphics Group for many useful discussions and suggestions.

This work was supported in part by grants from Apple, DEC, Hewlett Packard, and IBM. Additional support was provided by NSF (ASC-89-20219), as part of the NSF/DARPA STC for Computer Graphics and Scientific Visualization. All opinions, findings, conclusions, or recommendations expressed in this document are those of the author and do not necessarily reflect the views of the sponsoring agencies.

References

[dV70] Leonardo da Vinci. *Notebooks I, II*. Dover, New York, 1970.

[Gla94] G. Glaeser. *Fast Algorithms for 3-D Graphics*. Springer-Verlag, New York, 1994.

[Hag76] Margaret A. Hagen. Influence of picture surface and station point on the ability to compensate for oblique view in pictorial perception. *Developmental Psychology*, 12(1):57–63, January 1976.

[Hag80] Margaret A. Hagen, editor. *The Perception of pictures*. Academic Press series in cognition and perception. Academic Press, New York, 1980.

[HEJ78] Margaret A. Hagen, Harry B. Elliott, and Rebecca K. Jones. A distinctive characteristic of pictorial perception: The zoom effect. *Perception*, 7(6):625–633, 1978.

[Kub86] Michael. Kubovy. *The psychology of perspective and Renaissance art*. Cambridge University Press, Cambridge <Cambridgeshire> ; New York, 1986.

[PFTV88] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Cambridge, UK, 1988.

[Pir70] M. H. Pirenne. *Optics, painting and photography*. Cambridge University Press, New York, 1970.

[Pon62] L. S. Pontriagin. *The mathematical theory of optimal processes*. Interscience Publishers, New York, 1962.

[Rau86] B. V. Raushenbakh. *Sistemy perspektivy v izobrazitel'nom iskusstve : obshchaia teoriia perspektivy*. Nauka, Moskva, 1986. in Russian.

[TR93] Jack Tumblin and Holly E. Rushmeier. Tone reproduction for realistic images. *IEEE Computer Graphics and Applications*, 13(6):42–48, November 1993. also appeared as Tech. Report GIT-GVU-91-13, Graphics, Visualization & Usability Center, Coll. of Computing, Georgia Institute of Tech.

[Zor95] Denis Zorin. Perceptual distortions in images. Master's thesis, Caltech, 1995.

Figure 7. A wide-angle photograph of a room. a. Original image (approximately 100° angle. b.Transformation 1 applied with λ = 0. c. Generalization of transformation 1, (Section 5.4) applied, λ varies from 0 to 1.0 across the image. Note the correct shape of the head and straightness of the walls.

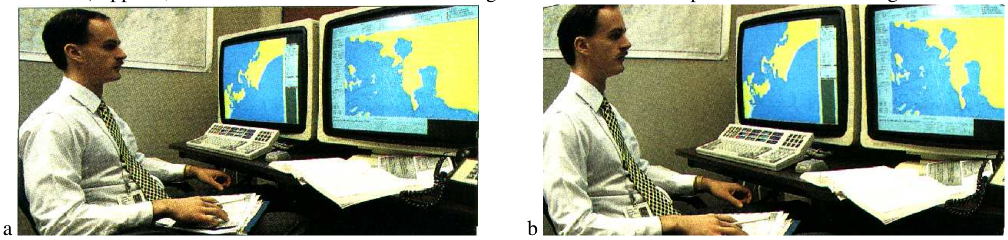

Figure 8. Photo from the article "Navigating Close to Shore" by Dave Dooling ("IEEE Spectrum", Dec. 1994), © 1994 IEEE, photo by Intergraph Corp. 92° viewing angle. a. Original image. b. Transformation 1 applied with λ = 0.

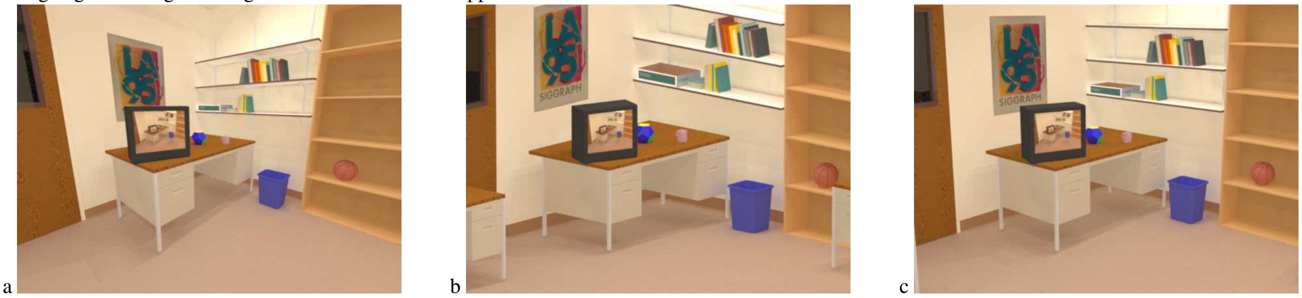

Figure 9. Shallow scene: model of an office (frames from the video), standard projections. a. 92° viewing angle. b. 3° viewing angle. c. 36° viewing angle, close to perceptually optimal for most people.

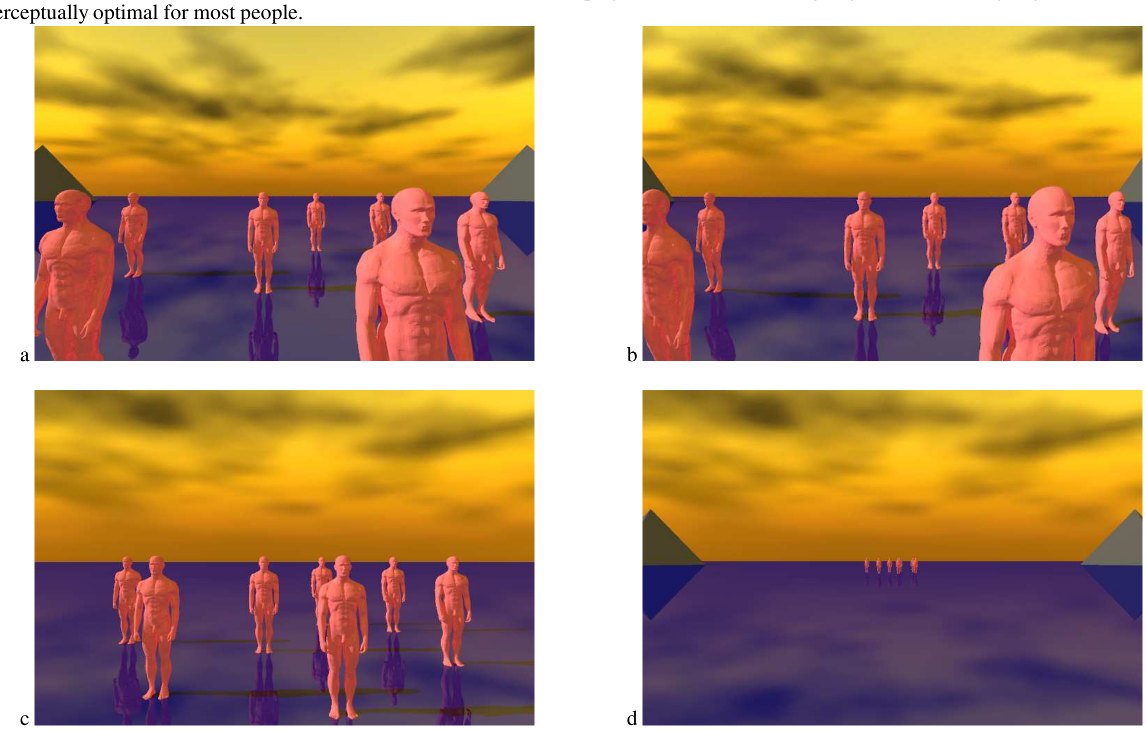

Figure 10. Deep scene, (frames from the video). If we want to have large images of men in the foreground while keeping the images of pyramids in the background, we have to make the angle of the picture wide enough. a. 90° viewing angle. Note the distorted form of the head of the men near the margins of the picture and differences in the shape of the bodies of the men in the middle and close to the margins. b. 90° viewing angle, transformation 1 applied, λ = 0. c. 60° viewing angle, keeping pyramids in the same position in the picture. d. 60° viewing angle, keeping people in the center in the same position.