Rendering Real-World Objects Using View Interpolation

Tomáš Werner Roger David Hersch* Václav Hlaváč

Department of Control Engineering
Faculty of Electrical Engineering
Czech Technical University
Karlovo náměstí 13
Prague 12135, Czech Republic

*Peripheral Systems Laboratory
École Polytechnique Fédérale de Lausanne
INF-Ecublens
Lausanne CH-1015, Switzerland

Abstract

This paper presents a new approach to rendering arbitrary views of real-world 3-D objects of complex shapes. We propose to represent an object by a sparse set of corresponding 2-D views, and to construct any other view as a combination of these reference views. We show that this combination can be linear, assuming proximity of the views, and we suggest how the visibility of constructed points can be determined. Our approach makes it possible to avoid difficult 3-D reconstruction, assuming only rendering is required. Moreover, almost no calibration of views is needed. We present preliminary results on real objects, indicating that the approach is feasible.

1 Introduction

Methods which are able to capture a real object and render it from an arbitrary viewpoint usually use a 3-D model of the object. The bottleneck of these methods is the 3-D reconstruction, which is a non-trivial problem, often failing for objects of more complex shapes. In our work, we show that, provided that only rendering is required, the object can be represented as a sparse set of views rather than a 3-D model, and the model reconstruction can thus be avoided. The new bottleneck becomes the correspondence problem, being simpler than 3-D reconstruction. Thus, more complex objects can be handled. In addition, faster access to a view can be achieved than by rendering the 3-D model.

As far as we know, there are very few papers on using view interpolation for accessing a view of a 3-D object from an arbitrary viewpoint. The most relevant is the recent paper [6] due to Laveau and Faugeras, in which the authors propose to represent a 3-D scene by a collection of corresponding reference views and fundamental matrices. They make use of the fact that the position of a point in a new view can be found as an intersection of the epipolar lines associated with the corresponding points in a pair of different views [2]. The visibility of this new point is determined by a ray-tracing-like algorithm. The apparent drawback of this approach is that the positions of new points are very sensitive to errors in calibration.

More frequent are papers dealing with prediction of positions of new pixels, though authors usually do not mention the possibility of representing the whole 3-D object by a set of 2-D views. The work of Ullman and Basri [10] describes an approach to 3-D object recognition using a set of views on an object instead of its model. It is shown that any instance of the object can be expressed as a linear combination of reference views assuming affine camera. A similar and even earlier work is presented in [5].

Quite general survey of so-called algebraic function of views can be found in [8]. These functions express the relations of projections of a single 3-D point in different views. We will mention basic results of this area later in the paper.

Our approach involves the representation of the object by a set of reference views, rather than by its opaque 3-D model. Any view not contained in this set is obtained as a combination of a small number of the closest reference views. We show that it can be computed as a *linear combination* of a subset of the views under an assumption of their proximity. We suggest how to solve the visibility problem in the new view. The way of referring the views also allows us to almost avoid calibration of the views (possibly except for the correspondence acquisition).

The automated 3-D reconstruction usually requires reliable and accurate range maps, obtained by an active range finder. The robustness of our alternative approach makes it possible to use only a *passive* stereo matcher. In contrast to 3-D reconstruction, view interpolation is more insensitive to those errors in the correspondence, which are caused by ambiguity in the areas of constant intensity. These errors are inevitable when a passive matching algorithm is used.

The presented work is a part of the project *Interview*, aiming to develop a prototype of a device which is able to capture and render complex real-world objects using view interpolation. The rendering algorithm should be implemented on a multiprocessor system. We expect to achieve a frame generation rate sufficient for displaying moving objects on the screen in real-time.

2 Our approach

Let us outline how to render an object from an arbitrary viewpoint without having its 3-D model. A set of reference views covering the whole visible surface of

Proceedings of the Fifth International Conference on Computer Vision (ICCV '95)
0-8186-7042-8/95 $10.00 © 1995 IEEE

the object is captured. These views can be accessed directly. What is demanded are the intermediate views. If the correspondence of the reference views is available, it is possible to obtain any interpolated view as a composition of a subset of the reference views close to it.

To succeed, the following problems must be solved:

1. How to predict the position and the intensity of a point in the new view if the positions and the intensities of corresponding points in the reference views are known?

2. How to choose the subset of reference views from which the required new view can be optimally constructed?

3. How to determine the visibility of points in the new view?

4. How to find the optimal set of necessary reference views?

5. How to find the correspondences between reference views?

In the following sections, we propose solution to some of these problems, namely to 1, 2, 3 and 5. As for 4, the question of the choice of the smallest and still sufficient set of the reference views is non-trivial. The more restricted fundamental problem of how to choose a set of so-called *characteristic views* (i.e., the minimum set of views in which all points of a given surface are visible, e.g. [5]) still remains unsolved for general non-convex objects. In our case, the set of reference views would have to meet additional requirements, e.g., a reasonable sampling frequency for all places on the surface. We do not deal with question 4 in this paper.

2.1 Possible approaches to prediction of positions of new points

Let us assume we know image coordinates, $\mathbf{x}_1, \ldots, \mathbf{x}_n$, of the projections of a 3-D scene point $\mathbf{X}$ in n different views. What is the smallest n which allows to determine image coordinates $\mathbf{x}$ of projection of $\mathbf{X}$ in a new, different view? If the views are *fully calibrated*, three image coordinates (e.g., two from one view, one from the other, that means $n = 2$) are sufficient to recover $\mathbf{X}$ and thus also its projection, $\mathbf{x}$, in the new view.

If the views are *weakly calibrated*, again three coordinates are sufficient to recover $\mathbf{x}$ (but no more $\mathbf{X}$). In this case, $\mathbf{x}$ is obtained as an intersection of epipolar lines associated with $\mathbf{x}_1$ and $\mathbf{x}_2$ [2]. Here the problem occurs if $\mathbf{x}$ lies on the intersection of the projection plane of the new view and the trifocal plane, i.e., the two epipolars are parallel. In fact, if the epipolars are nearly parallel, $\mathbf{x}$ cannot be determined in practice [6].

The alternative approach is to use *algebraic functions of views* [8], which are algebraic relations among image coordinates of projections of a single scene point in different views (e.g., the epipolar constraint is an algebraic function of two views). Again, $n = 2$ is sufficient. For perspective views, it can be shown that each component of $\mathbf{x}$ can be expressed as a *ratio of bilinear functions* of the components of $\mathbf{x}_1$ and $\mathbf{x}_2$. For orthographic views, each component of $\mathbf{x}$ is a *linear function* of components of $\mathbf{x}_1$ and $\mathbf{x}_2$.

2.2 Parametrization of the views

We propose that the reference views and the new view are referred to not by their camera parameters or fundamental matrices but rather by a certain parameter $\mathbf{p}$, whose relation to the real camera parameters need not be known. Let us denote by $\mathcal{I}_{ref}$ the set of reference views $I_1, I_2, \ldots$. The view I_i is thus referred to by a parameter $\mathbf{p}_i$. Parameter $\mathbf{p}$ parametrizes the set $\mathcal{P}$ of allowed viewpoints and orientations of the camera. We assume $\mathbf{p}$ to be a continuous function of camera parameters such that "near" views have near $\mathbf{p}$. For example, if we wish to display an object from any viewpoint placed on a viewing sphere, we can choose $\mathbf{p}$ to be a pair of angles of spherical coordinates of points on the viewing sphere. Another example is an uncalibrated time sequence, where $\mathbf{p}$ can be the time.

A new view, I, is accessed by its parameter, $\mathbf{p}$. I can be obtained as a combination[1] of a subset $\mathcal{I}^{\mathrm{p}}_{ref} \subset \mathcal{I}_{ref}$ of reference views that are close to I:

$$I = g(\mathcal{I}^{\mathrm{p}}_{ref}, \mathbf{a}(\mathbf{p})) \tag{1}$$

$\mathbf{a}(\mathbf{p})$ is a vector of parameters of the combination. As we do not know how $\mathbf{p}$ is related to real camera parameters, we cannot determine positions of the points in I exactly. We must find such g, $\mathcal{I}^{\mathrm{p}}_{ref}$ and $\mathbf{a}(\mathbf{p})$ (the latter two vary with $\mathbf{p}$), so that:

1. If $\mathbf{p} = \mathbf{p}_i$ for some i, then $g(\mathcal{I}^{\mathrm{p}}_{ref}, \mathbf{a}(\mathbf{p})) = I_i$.

2. Otherwise, $g(\mathcal{I}^{\mathrm{p}}_{ref}, \mathbf{a}(\mathbf{p})) = I_i$ is a good approximation of a real view with parameter $\mathbf{p}$.

One of the advantages of this approach is that even if errors (e.g., due to mismatches in the correspondence) can occur in I when 2 holds, it is ensured that I is correct when 1 holds. This contributes a great deal to the robustness of the method and makes it possible to choose a trade-off between the number of the reference views and the fidelity of the displayed views. Moreover, no camera calibration is needed up to this moment because instead of the camera parameters, the parameter $\mathbf{p}$ is used. The penalty is the necessity to approximate.

2.3 Linear interpolation among near views

We will show that a quite simple choice of g, $\mathcal{I}^{\mathrm{p}}_{ref}$ and $\mathbf{a}(\mathbf{p})$ is sufficient to predict positions of points in the new view.

The image coordinates $\mathbf{x}$ of a projection of a 3-D scene point in a view with parameter $\mathbf{p}$ are given as a non-linear function:

$$\mathbf{x} = f(\mathbf{X}, \mathbf{p}) \tag{2}$$

Let us express the parameter $\mathbf{p}$ of the new view as a linear combination of parameters of n reference views,

[1] By *combination* is meant a function of image coordinates of corresponding points.

Proceedings of the Fifth International Conference on Computer Vision (ICCV '95)
0-8186-7042-8/95 $10.00 © 1995 IEEE

$\mathbf{p} = \sum a_i \mathbf{p}_i$, assuming also that $\sum a_i = 1$. If the parameters $\mathbf{p}_i$ are close enough to each other, we obtain:

$$f(\mathbf{X}, \sum_{i=1}^{n} a_i \mathbf{p}_i) \approx \sum_{i=1}^{n} a_i f(\mathbf{X}, \mathbf{p}_i), \tag{3}$$

i.e.,

$$\mathbf{x} = \sum_{i=1}^{n} a_i \mathbf{x}_i \tag{4}$$

The proof can be made by substituting $\mathbf{p}_i = \mathbf{q}_0 + \Delta \mathbf{q}_i$, $\mathbf{p} = \mathbf{q}_0 + \sum a_i \Delta \mathbf{q}_i$, and by expanding the function $h(\mathbf{p}) = f(\mathbf{X}, \mathbf{p})$ to its Taylor series along the point $\mathbf{q}_0$, considering only the first order term.

The numbers a_i can be obtained by solving the equations:

$$\begin{aligned} \sum_{i=1}^{n} a_i \mathbf{p}_i &= \mathbf{p},\ i = 1, \ldots, n \\ \sum_{i=1}^{n} a_i &= 1 \end{aligned} \tag{5}$$

If $\mathbf{p}_i$, $i = 1, \ldots, n$, are linearly independent, the system (5) has a unique solution for $n = \dim(\mathbf{p}) + 1$. It implies that we need $\dim(\mathbf{p}) + 1$ reference views to determine the new one.

Thus we have[2] $g(\mathcal{I}^{\mathrm{p}}_{ref}, \mathbf{a}(\mathbf{p})) = \sum_{i=1}^{n} a_i \mathbf{x}_i$ and $\mathbf{a}(\mathbf{p}) = (a_1, \ldots, a_n)$. For a certain $\mathbf{p}$, $\mathcal{I}^{\mathrm{p}}_{ref}$ can be chosen as $n = \dim(\mathbf{p}) + 1$ such reference views, which have linearly independent parameters and which are close enough to $\mathbf{p}$. In fact, $\mathbf{x}$ is obtained using n-dimensional linear interpolation.

The same procedure can be used for interpolation of image intensities (or values of color components) if we formally replace f with a reflectance (or color) model.

2.4 Algebraic functions of views

As a more accurate prediction of points in the new view, the algebraic functions of views (see section 2.1) can be used. E.g., assuming orthographic views, the following relations hold:

$$\begin{aligned} x^1 &= \alpha_{11} x_1^1 + \alpha_{12} x_2^1 + \alpha_{13} x_2^2 + \alpha_{14}, \\ x^2 &= \alpha_{21} x_1^2 + \alpha_{22} x_2^1 + \alpha_{23} x_2^2 + \alpha_{24}, \end{aligned} \tag{6}$$

where $\mathbf{x} = (x^1, x^2)$ are again image coordinates of the new point and $\mathbf{x}_1 = (x_1^1, x_1^2)$ and $\mathbf{x}_2 = (x_2^1, x_2^2)$ are image coordinates of the corresponding points from a chosen pair of reference views. The parameters of the combination are $\mathbf{a}(\mathbf{p}) = (\alpha_{kl})$.

Each $\mathcal{I}^{\mathrm{p}}_{ref}$ uniquely determines its "region of influence" as the set of such values of $\mathbf{p}$ for which $\mathcal{I}^{\mathrm{p}}_{ref}$ stays constant. Let us denote this set $\mathcal{P}^{\mathrm{p}}$. For each $\mathcal{I}^{\mathrm{p}}_{ref}$, the parameters $\mathbf{a}(\mathbf{p})$ must be found for all $\mathbf{p} \in \mathcal{P}^{\mathrm{p}}$. This can be done by interpolation with the independent variable $\mathbf{p}$ and the dependent variable $\mathbf{a}(\mathbf{p})$, using pre-computed values of $\mathbf{a}(\mathbf{p}_i)$ for several parameters, $\mathbf{p}_i$, of appropriately chosen reference views, I_i. These reference views I_i are chosen so that $\mathbf{p}_i$ are near the region $\mathcal{P}^{\mathrm{p}}$.

Each parameter $\mathbf{a}(\mathbf{p}_i)$ can be computed as a solution of a (possibly overdetermined) linear equation system. This system is obtained by substituting a sufficient number of corresponding point triplets to the system (6). These triplets, $[\mathbf{x}, \mathbf{x}_1, \mathbf{x}_2]$, are such that $\mathbf{x} \in I_i$, $\mathbf{x}_1 \in I_r$, $\mathbf{x}_2 \in I_s$ and $I_r, I_s \in \mathcal{I}^{\mathrm{p}}_{ref}$.

2.5 Visibility of interpolated points

In the sections above, we dealt with a prediction of image coordinates of a single point in the new view. The whole view can be constructed by repeating this prediction for all corresponding n-tuples of points of the subset $\mathcal{I}^{\mathrm{p}}_{ref}$.

We will make a restriction to opaque objects in this paper. Let us denote p_i and p the projections of a point X of the scene in views I_i and I with parameters $\mathbf{p}_i$ and $\mathbf{p}$, respectively ($I_i \in \mathcal{I}^{\mathrm{p}}_{ref}$). Let us now distinguish the six possible cases that can arise for visibility of X in particular views (Fig. 1) and discuss how they influence the construction of the new view, I:

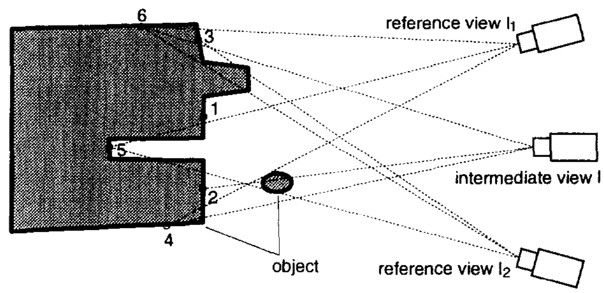

Figure 1: The six cases which can arise for the visibility of points in the interpolated view. The numbers correspond to the notation in text.

1. All p_i are visible, p is visible.
   The interpolated point p can be constructed from $\mathcal{I}^{\mathrm{p}}_{ref}$ and it is visible. This case should occur for as many points of I as possible.
2. All p_i are visible, p is invisible.
   The point X is occluded in I by some other scene point Y, but the interpolated point p may be constructed from the given subset. This case occurs if for some other n-tuple of corresponding points the combination $g(\mathcal{I}^{\mathrm{p}}_{ref}, \mathbf{a}(\mathbf{p}))$ gives the same result (and it can be also detected like this). It is necessary to decide which of the two points X and Y is closer to the projection plane of I. We will discuss this case later.

[2] This is a bit formal because g was defined as a function of a set of views and here we have a function of image coordinates on the right-hand side. For the sake of simplicity, g will denote both the function of a set of views, $g(\mathcal{I}^{\mathrm{p}}_{ref}, \mathbf{a}(\mathbf{p}))$, and the function of an n-tuple of corresponding points, $g([\mathbf{x}_1, \ldots, \mathbf{x}_n], \mathbf{a}(\mathbf{p}))$.

Proceedings of the Fifth International Conference on Computer Vision (ICCV '95)
0-8186-7042-8/95 $10.00 © 1995 IEEE

3. Some p_i are visible and some invisible, p is visible. The point X is occluded in some reference view, I_i. Therefore the n-tuple of corresponding points, whose combination would yield the coordinates of p, does not exist. The areas where these situations occur remain empty[3] in the constructed view. The missing points can be obtained using a slightly different subset of reference views provided that X is not occluded in them.

4. Some p_i are visible and some invisible, p is invisible.
   The point X is not visible from the view I and it also cannot be constructed as a combination of any n-tuple of corresponding points from $\mathcal{I}^{\mathrm{P}}_{ref}$.

5. All p_i are invisible, p is visible.
   The handling of this situation is the same as in the case 3.

6. All p_i are invisible, p is invisible.
   The handling is the same as in the case 4.

We see it is necessary to detect and correctly handle the cases 2, 3, and 5.

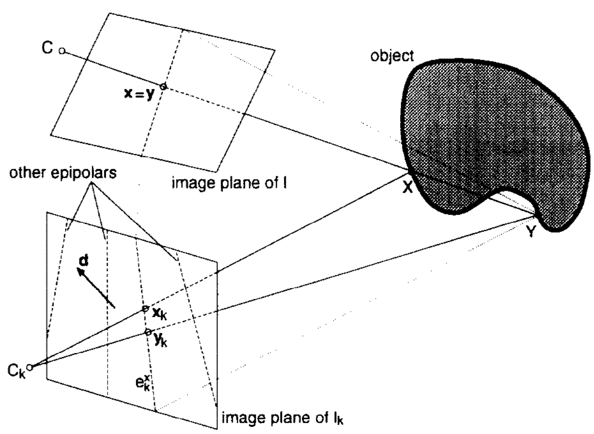

Figure 2: Determining which of the points, X and Y, is closer to the center of projection of I.

The case 2 requires to find out which of the two scene points, X and Y, whose projections have the same image coordinates in the new view I, is closer to the projection plane of I. There exist two n-tuples, $[\mathbf{x}_1, \ldots, \mathbf{x}_n]$ and $[\mathbf{y}_1, \ldots, \mathbf{y}_n]$, of corresponding points from $\mathcal{I}^{\mathrm{P}}_{ref}$ such that:

$$g([\mathbf{x}_1, \ldots, \mathbf{x}_n], \mathbf{a}(\mathbf{p})) = g([\mathbf{y}_1, \ldots, \mathbf{y}_n], \mathbf{a}(\mathbf{p})) \qquad (7)$$

Let us choose one of the reference views, $I_k \in \mathcal{I}^{\mathrm{P}}_{ref}$. Then $\mathbf{x}_k$ and $\mathbf{y}_k$ are two different points in I_k, lying on the epipolar line e^x_k associated with $\mathbf{x}$ (see Fig. 2). As e^x_k is the projection of the line connecting the center of projection of I, X, and Y, then mutual position of $\mathbf{x}_k$ and $\mathbf{y}_k$ on e^x_k reflects the order of depths of X and Y.

In fact, it is not necessary to know the exact epipolar geometry. It is enough to find a function which is proportional to the difference of the depths of X and Y in I. It can be chosen as:

$$d(\mathbf{x}_k, \mathbf{y}_k) = (\mathbf{x}_k - \mathbf{y}_k) \cdot \mathbf{d} \qquad (8)$$

The sign of $d(\mathbf{x}_k, \mathbf{y}_k)$ decides which of the two points, X and Y, is closer.

The vector $\mathbf{d}$ must express the approximate direction of epipolars. More exactly, for all points $\mathbf{z}_k \in I_k$, the inner product $(\mathbf{z}_k - \mathbf{e}_k) \cdot \mathbf{d}$ must have the same sign ($\mathbf{e}_k$ is the epipole of I_k). Assuming the epipole, $\mathbf{e}_k$, is outside I_k, $\mathbf{d}$ can be constant for all $\mathbf{z}_k$. Its choice is not critical — its direction can be determined much more easily than the fundamental matrix, using the point correspondences from I and I_k. Yet it is an open question if its orientation can also be found using only these correspondences. If not, $\mathbf{d}$ must be known *a priori* — this is the only calibration needed.

The cases 3 and 5 require that all points of the object's surface are visible at least in n reference views. This must be ensured by a proper selection of $\mathcal{I}_{ref}$.

2.6 Correspondence acquisition

We need to find the correspondence for the subsets of the reference views from which the new views are to be constructed. In other words, we are looking for all n-tuples $[\mathbf{x}_1, \ldots, \mathbf{x}_n]$ of pixels coordinates, each from a different view, so that every n-tuple contains projections of a single point on the object's surface, without considering errors caused by discretization and noise. This *correspondence problem* is difficult mainly due to (1) the change of geometric and photometric properties with the viewpoint, (2) the presence of noise, (3) discretization, and (4) the lack of information in areas of constant intensity.

Rendering an object by combining reference views brings an important advantage over the approaches using a 3-D model — namely low sensitivity to correspondence errors in areas of (almost) constant intensity. The areas in which these errors appear in the constructed view will also have a constant intensity, so the image will *look* the same. Even if the visibility of the interpolated pixel is inferred from the correspondence (the case 3 in the last section about visibility), only minor artifacts can be expected because no incorrect facets are constructed as it would be the case with the 3-D model. This insensitivity allows us to use passive methods for acquiring the correspondence, having the potential to capture both indoor and outdoor objects, and the possibility to provide directly the dense correspondence.

There are two alternatives for determining the correspondence of n views: (i) n-ocular stereo, and (ii) composition of pairwise correspondences[4]. For now we

[3] These areas must be distinguished from the holes caused by unequal sampling frequencies in the views. These holes can be filled by, e.g., interpolation using intensities of neighboring pixels.

[4] Even if the accumulation of errors of pairwise correspondences limits the number of compositions, this composition is feasible for small n.

Proceedings of the Fifth International Conference on Computer Vision (ICCV '95)
0-8186-7042-8/95 $10.00 © 1995 IEEE

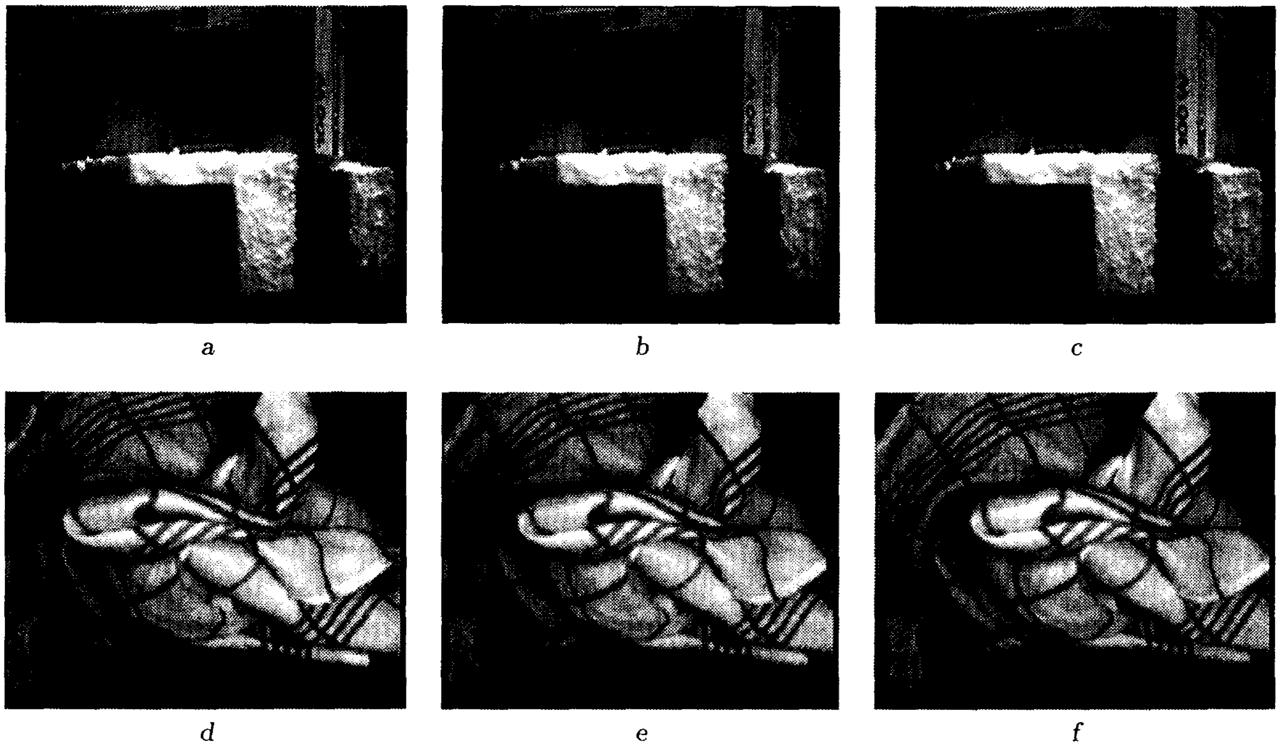

Figure 3: Two reference views of the first object (consisting of a book, a piece of styrofoam, a box and a match box) , $\mathbf{p}_1 = [0°], \mathbf{p}_2 = [5°]$ (a, c) and the predicted view of the first object, $\mathbf{p} = [2.5°]$, $a_1 = a_2 = 0.5$ (b). Two reference views of the second object (a linen towel), $\mathbf{p}_1 = [0°], \mathbf{p}_2 = [10°]$ (d, f) and the predicted view of the second object $\mathbf{p}_1 = [5°]$, $a_1 = a_2 = 0.5$ (e).

chose the alternative (ii) because the binocular stereo matcher is easier to implement than the n-ocular one but it is sufficient for our experiments.

We implemented a modified version of the binocular stereo matcher [1]. It is similar to the more known algorithm [7], but rather than intervals between edges, the raw pixel intensities are matched. Epipolar, uniqueness, and ordering constraints are utilized. For each pair of corresponding epipolars, a pair of non-decreasing transformation functions is found. This pair of functions transforms the intensity functions so that the sum of costs of matches and occlusions is minimized. This optimization problem is solved by dynamic programming. The cost of a match and occlusion is derived using the Bayesian approach. The advantage of the algorithm is that it directly produces the dense correspondence.

3 Experiments

We made several experiments with the construction of intermediate views. We used real objects of quite complex shapes, for which the 3-D reconstruction would be difficult. For simplicity, objects were allowed only to rotate around a single axis, the view parameters $\mathbf{p}$ thus having one component (the angle).

For prediction of both the image coordinates and the intensities of pixels in new views, the linear interpolation (4) was used. Since $\dim(\mathbf{p}) = 1$, two reference views are needed for determining interpolated views.

Our experimental setup consisted of a camera, a calibrated turntable, and a calibration grid. The calibration was necessary only to find epipolar geometry for the stereo matcher. The objects were placed on the table, and the views changed by rotating the table. Two reference views of each object were captured and their correspondence was found using the described stereo matcher. Then new views were constructed.

The upper three images in Fig. 3 show the result for the first object. The difference between azimuth angles was 5°. Occluded areas, as well as holes due to unequal sampling frequencies were filled in by linear interpolation of intensities on epipolars, without considering the solutions described in section 2.5. Even so, no artifacts in the interpolated view are visible.

The lower three images in Fig. 3 show a more complex object. The difference of the azimuth angles is 10°. As large areas of the object's surface are occluded, pixels in or near these areas were matched incorrectly by the stereo matcher (see, e.g., the stripe in the middle-top of the image e). While this would cause rough errors in the 3-D model, the appearance of the new view is still acceptable here.

Proceedings of the Fifth International Conference on Computer Vision (ICCV '95)
0-8186-7042-8/95 $10.00 © 1995 IEEE

4 Conclusion

This paper is an attempt to show that a 3-D model is not the only possible representation of a 3-D object, provided that only rendering is required. We have proposed an alternative representation by a sparse set of captured reference views. The missing views are determined using information about their mutual correspondence. Furthermore, we have dealt with construction of these new views. We have shown that the interpolated view can be expressed as a linear combination of the reference ones, assuming the proximity of the views, and how the visibility of points in the interpolated view can be determined. We have not dealt with the important question of the the optimal set of reference views.

The experiments have proven the possibility to construct intermediate views for quite complicated objects. The presented results indicate the feasibility and robustness of our approach.

Our method should be an alternative to existing approaches based on 3-D model reconstruction. We do not want to eliminate the need for 3-D models but rather to give a possibility to choose a trade-off between the number of reference views (and thus the volume of data necessary for representing the object) on the one side, and fidelity of "appearance" of rendered views and difficulty of the algorithm on the other side. In fact, our method is specially optimized for rendition purposes. Therefore, e.g. high geometric or photometric accuracy can be sacrificed in favor of approximative relations.

In the future, we plan to make experiments with predicting positions of new views using algebraic functions of views and with the proposed algorithms for determining the visibility of the predicted points. We also need to improve the correspondence acquisition process. Finding at least a partial solution to the problem of the optimal set of reference views could be a topic for an independent research. Finally, we intend to implement the rendering algorithm on a multiprocessor-multidisk system [4]. We expect that it will enable displaying moving objects in real-time.

Acknowledgement

The project is supported by the Swiss National Fund (grant No. 83H-036863), by the Grant Agency of the Czech Republic (grant No. 102/93/0954), and by the Czech Technical University (grant No. 38196).

We are indebted to all the people whose comments contributed to the quality of the paper, but mainly to Dr. Radim Šára from the Computer Vision Lab, CTU Prague, for his careful reading of the whole text.

References

[1] I. J. Cox, S. Hingorani, B. M. Maggs, and S. B. Rao. Stereo without disparity gradient smoothing: a Bayesian sensor fusion solution. In *British Machine Vision Conference*, pages 337–346, Berlin, 1992. Springer-Verlag.

[2] O. Faugeras and L. Robert. What can two images tell us about a third one? In *European Conference on Computer Vision*, pages 485–492, 1994. Also the technical report 1993, available on ftp.inria.fr.

[3] O. D. Faugeras, Q. T. Luong, and S. J. Maybank. Camera self-calibration: Theory and experiments. In *European Conference on Computer Vision*, pages 321–333, 1992.

[4] B. A. Gennart, B. Krummenacher, L. Landron, and R. D. Hersch. Gigaview parallel image server performance analysis. In *World Transputer Congress'94*, pages 120–135. IOS Press, September 1994.

[5] I. Chakravarty and H. Freeman. Characteristic views as a basis for three-dimensional object recognition. In *Proc. of the Society for Photo-Optical Instrumentation Engineers Conference on Robot Vision*, pages 37–45, 1982.

[6] S. Laveau and O. Faugeras. 3-D scene representation as a collection of images. In *Proc. of 12th International Conf. on Pattern Recognition, Jerusalem, Israel*, pages 689–691, October 9–13 1994. Also the technical report, available on ftp.inria.fr.

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

[8] A. Shashua. On geometric and algebraic aspects of 3D affine and projective structures from perspective 2D views. Technical Report AI Memo No. 1405, C.B.C.L. Paper No. 78, Massachusetts Institute of Technology, Artifical Intelligence Laboratory, July 1993.

[9] R. Skerjanc and J. Liu. Computation of intermediate views for 3DTV. In R. Klette and W. G. Kropatsch, editors, *Theoretical Foundations of Computer Vision*. Akademie Verlag, 1992.

[10] S. Ullman and R. Basri. Recognition by linear combinations of models. Memo 1152, MIT Artificial Intelligence Laboratory, August 1989.

[11] A. Zakhor and F. Lari. 3D camera motion estimation with applications to video compression and scene reconstruction. In *SPIE Symposium on Electronic Imaging*, pages 1–14, Dept. of Electrical Eng. and Computer Sciences, University of California, Berkeley CA94720, 1993.

Proceedings of the Fifth International Conference on Computer Vision (ICCV '95)
0-8186-7042-8/95 $10.00 © 1995 IEEE