Displaying 3D Real-Objects Using 2D View Extrapolation for Virtual Museums

Václav Hlaváč, Tomáš Werner, Tomáš Pajdla
Czech Technical University, Faculty of Electrical Engineering
Center for Machine Perception*
121 35 Prague 2, Karlovo namĕstí 13, Czech Republic
{hlavac,werner,pajdla}@vision.felk.cvut.cz

1 Task formulation

This contribution informs about the recent computer vision technology for displaying an 3D object from any viewpoint on computer screen without creating its full 3D model. We believe that this our approach has high potential in museums.

The technology is the following. The object of interest for a virtual museum visitor (e.g. a small sculpture, an old jewel) is placed on the turntable. A set of 2D intensity or colour images that covers the range of needed views is captured. Then the computational procedure—to be sketched in this paper—is run. The outcome is the object representation that allows the user to look at the object on the computer screen from varying viewing directions. She or he can interactively control from where he looks. This approach can run remotely via Internet and thus fits into the concept of a virtual museum.

This is just one application of more general approach called *telepresence*. It allows to give the remote person the feeling he is in a different place. Let us demonstrate this idea on application in medicine. An expert can give advices to the operating team from a distant place. The input is several dozens of cameras placed around the patient in the remote operating room.

Authors of this contribution come from a university research team. They are interested in the research issues to the approach and had some success in this respect. The purpose of this contribution is to inform people from the museums about the recent and relatively simple technology. We believe that such systems will be on the market very soon.

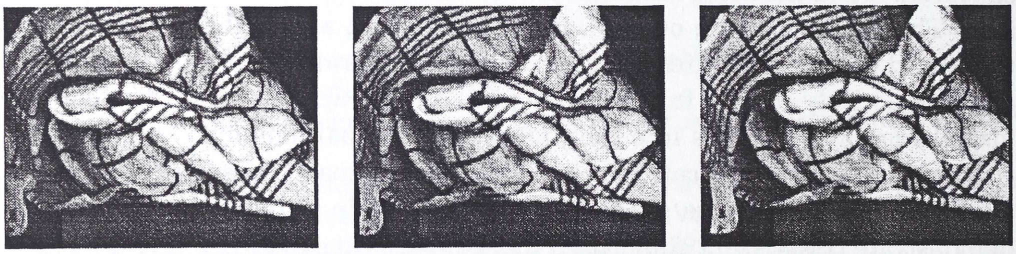
*This research was supported by the Czech Ministry of Education grant VS96049, the Grant agency of the Czech Republic grant 102/97/0855, the EU Copernicus grant CP941068.

2 Images instead of a traditional rendering of a 3D model

Methods which are able to capture a real object and render it from an arbitrary viewpoint usually use a 3D geometric model of the object. The computer graphics has attacked this problem for three decades with considerable success. The bottleneck of these methods is the 3D reconstruction of a model, which is a non-trivial problem, often failing for objects of more complex shapes. Therefore the approach is not used in museums.

Alternatively, *image-based scene representation paradigm* proposes to display a real 3D scene from any viewing direction without using its 3D model. The scene is represented by a collection of 2D *reference views* instead of a full 3D model. The actual image to be displayed is called a *virtual view* and is created by interpolation from the reference views using correspondences among them. The new bottleneck becomes the correspondence problem, being simpler than 3D reconstruction. The aim of such procedure is to avoid the difficult problem of consistent 3D model reconstruction. Thus, more complex objects can be handled. In addition, faster access to a view can be achieved than by rendering the 3D model.

The image-based approach copes with complicated free-form surfaces as is demonstrated in Fig. 1. Notice the mistake in the interpolated image in the middle. One of the lines in the top, centre is doubled. This is due to mismatch in correspondences. This error is likely to be overlooked. This shows how human understanding of the image content is not sensitive to such errors.



(a) (b) (c)

Figure 1: Left and right are two reference views of a linen towel. View directions are 10° apart. The virtual image in the middle was obtained using interpolation.

3 The partial tasks to be solved

To succeed, the following problems must be solved:

Correspondence problem. How to find the correspondences between almost all points reference views? This issue is a difficult evergreen in computer vision. Our case allows when we can put the object (e.g. a small sculpture) on the turn table under our control makes the task much easier.

We proposed to capture dense sequence of images. The correspondence problem becomes almost trivial as we can track the important features on the object surface using a modification of an edge detector. Its input is an one-dimensional signal only as the epipolar geometry constraints the task.

View selection. How to find the optimal set of necessary reference views? We proposed a solution in [WHLP96]. The natural criterion of optimality is minimal number of views allowing synthesis of an image that looks similar as that would be seen when looking from the same viewpoint. It is not an easy task to come with a good measure of image similarity that will be close to human understanding of image similarity.

Image interpolation/extrapolation and geometry. 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?

It was thought [LF94, SD95, WHH95] that displaying a 3D scene from stored 2D images is quite different from rendering a 3D model. The difference seemed to follow from the observation of Ullman [UB91] who have proposed that the objects could be recognized just on the basis of linear combination of corresponding points in their orthographic images. We shall show in the sequel that knowledge of geometry is needed.

4 Knowledge of depth is needed to display properly

Ullman's approach has attracted new attention since Shasua showed that a trilinear function replaces the linear one [Sha93] for a perspective camera; since Laveau and Faugeras [LF94] and Hartley [Har95] have made clear that for the visualization itself any projective reconstruction of the scene suffices. Tedious calibration of the camera has been thus avoided in the case of visualization.

Seitz and Dyer [SD95] have stressed that visualizing an object by interpolating close views is a well posed process and therefore perfect correspondence algorithm are not ultimately needed for certain limited tasks. Other works have demonstrated that even quite complicated scenes can be shown by interpolating between the reference views [WHH95].

Yet, to display the scene by a camera revolving around it on a view circle, quite many reference images were needed in [WHH95] to make visual effect enough realistic. This is caused by the principal deficiency of image interpolation, namely in its *inability to show a general object from an arbitrary viewing angle using the images and the correspondences obtained from a sparse set of views*. Surprisingly, no object, not even a convex polyhedron, can be completely viewed by interpolating between finite number of reference views.

Consider the situation when reference views are located around a simple polyhedron as illustrated in Fig. 2. The images from the virtual camera C lying in the segment B_1C_4 cannot be constructed by interpolating reference views C_2 and C_3 since the camera C_2 does not see both sides of the polyhedron which are seen by the camera C. It will not help to move one of the reference cameras,

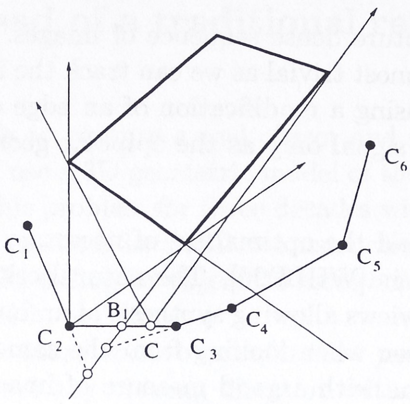


Figure 2: Virtual view C cannot be constructed by interpolation from the views C_2 and C_4 but can be extrapolated from C_3 and C_4.

e.g. C_3, closer to B_1 in hope to avoid malfunction of the view synthesis. By moving C_3 to B_1, the same problem appears on the segment C_3C_4. The only solution would be to increase the density of views near B_1 to infinity. Indeed, the algorithm for finding the best sparse set of reference views [WHLP96] has tended to select many reference views in places where aspect changed. Point B_1 is such a place in Fig. 2.

The above argument shows that the views cannot be constructed by interpolating the reference views from different aspects. On the other hand, it is certainly possible to construct view C from the views C_3 and C_4 by the *extrapolation of views.* Deep difference between the interpolation and the extrapolation manifests itself on the border between aspects where virtual camera has to switch from the views C_1, C_2 to the views C_3, C_4. Unlike in the interpolation case, where switching has been done at the centers of the reference cameras, here it is not clear where to switch between the reference views until the aspect of the object is not known. But finding the changes in aspect is equivalent to *finding depth discontinuities.* Moreover, for the reference views in general positions, it is not possible to move smoothly along the object just by linear extrapolation and switching on the borders between aspects as the line C_1C_2 need not to intersect the line C_3C_4 on the change of the aspect. When crossing the boundary of an aspect, it is also necessary to solve the *visibility of the points* because not all the points have to disappear at the same time.

Finding the discontinuities and resolving the visibility are problems known from 3D surface reconstruction and visualization of a 3D model. In order to synthesize the images from a virtual camera smoothly moving around the object, one has to step back from pure image based scene representation and interpolation mechanism to a partial projective reconstructions of the shape and their correct visualization.

5 Proposed technology and experiment

We developed a method for constructing virtual views outside of the aspect of the reference views [WPH97]. A *practically usable technology* that allows to build an image-based representation of a 3D scene routinely without human aid is proposed. The novel contribution is in the following topics:

- The key step towards view extrapolation is to solve the *correspondence problem.* The solution we propose is to employ *tracking of edge features in a dense sequence of images.*
- The *visibility of points in virtual views* is an important issue. Oriented projective geometry is used to formulate and solve the problem.
- It is shown that it is not needed to transfer each point from reference views when creating the virtual image. Instead, we propose *to triangulate the correspondences in reference images, to transfer exactly only their vertices, and to fill triangle interiors.*

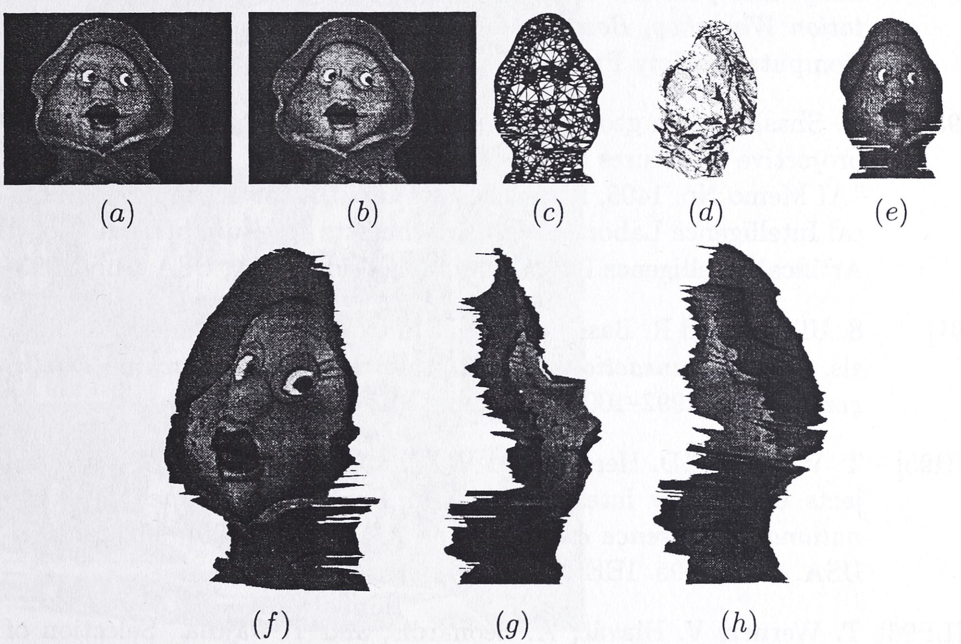


Figure 3: The results for the second object. (*a*) and (*b*) show the reference views, (*c*) the triangulation in the first reference view, (*d*) the projective model. (*e*) shows the virtual view from the same viewpoint as reference view, (*f*) and (*g*) other virtual views. (*h*) shows the incorrectly solved visibility in (*g*).

Let us demonstrate the displaying of a 3D object on a a ceramic doll, are in Fig. 3. The input are just two reference images that are captured from two view directions 10^o apart.

The Fig. 3e shows a virtual view from the same direction as the one of the reference views. It can be seen that the rendering of the triangulated surface gives very similar results. Fig. 3f, g demonstrates that virtual views can be extrapolated quite far from the couple of original reference views. The limit of this range of views is the viewing directions where there is not enough information in reference images to solve the visibility. This is demonstrated in Fig. 3h.

References

[Har95] R. I. Hartley. A linear method for reconstruction from lines and points. In *Proceedings of the 5th International Conference on Computer Vision*, pages 882–887, Boston USA, June 1995. IEEE Computer Society Press.

[LF94] 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.

[SD95] S. M. Seitz and C. R. Dyer. Physically-valid view synthesis by image interpolation. In *Proceedings of the Visual Scene Representation Workshop, Boston, MA., USA, June 24*, pages 18–27. IEEE Computer Society Press, 1995.

[Sha93] A. Shashua. On geometric and algebraic aspects of 3D affine and projective structures from perspective 2D views. Technical Report "AI Memo No. 1405, Massachusetts Institute of Technology, Artifical Intelligence Laboratory, Massachusetts Institute of Technology, Artifical Intelligence Laboratory, Cambridge, MA, USA, July 1993.

[UB91] S. Ullman and R. Basri. Recognition by linear combination of models. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 13(10):992–1005, October 1991.

[WHH95] T. Werner, R.D. Hersch, and V. Hlaváč. Rendering real-world objects using view interpolation. In *Proceedings of the 5th International Conference on Computer Vision*, pages 957–962, Boston, USA, June 1995. IEEE Computer Society Press.

[WHLP96] T. Werner, V. Hlaváč, A. Leonardis, and T. Pajdla. Selection of reference views for image-based representation. In *Proceedings of the 13th International Conference on Pattern Recognition*, volume I – Track A: Computer Vision, pages 73–77, Vienna, Austria, August 1996. IEEE Computer Society Press, Los Alamitos, CA., USA.

[WPH97] T. Werner, T. Pajdla, and V. Hlaváč. Visualizing 3-D real-world scenes using view extrapolation. Technical Report K335-CMP-1997-137, Czech Technical University, Dept. of Control, Faculty of Electrical Engineering, Czech Technical University, Karlovo náměstí 13, 12135 Praha, Czech Republic, June 1997.