Enabling View-Dependent Stereoscopic Projection in Real Environments

Oliver Bimber, Gordon Wetzstein, Andreas Emmerling and Christian Nitschke
Bauhaus-University Weimar
Bauhausstraße 11, 99423 Weimar, Germany,
{oliver.bimber, gordon.wetzstein, andreas.emmerling, christian.nitschke}@medien.uni-weimar.de

Abstract

We show how view-dependent image-based and geometric warping, radiometric compensation, and multi-focal projection enable a view-dependent stereoscopic visualization on ordinary (geometrically complex, colored and textured) surfaces within everyday environments. Special display configurations for immersive or semi-immersive AR/VR applications that require permanent and artificial projection canvases might become unnecessary. We demonstrate several ad-hoc visualization examples in a real architectural and museum application context.

1. Introduction and Motivation

Projector-based displays have clearly replaced head-attached displays for most virtual reality (VR) applications. Immersive surround screen displays and semi-immersive wall-like or table-like configurations are being used for visualizing two-dimensional or three-dimensional graphical content.

Today, the majority of augmented reality (AR) applications focuses on mobility. Thus wearable or portable devices have become dominant in this area. However, an increasing trend towards projector-based displays for AR can be noticed. A variety of stationary (e.g., [7,14,18], movable (e.g., [8,17]) and hand-held (e.g., [20]) projectors have been proposed for displaying graphical information directly on real objects or surfaces instead of performing optical augmentations or video compositions. These approaches, however, mimic real-world situations by applying simplified placeholders of real objects or environments that are optimized for projection (i.e., white surfaces with relatively basic 3D structure, such as multi-planar or parametric surfaces).

State-of-the-art multi-projector approaches are able to display seamlessly blended images onto geometrically and radiometrically complex surfaces [3,10]. Such techniques will open new possibilities for both – augmented reality and virtual reality. The immersive or semi-immersive visualization of stereoscopic three-dimensional graphics will become feasible within real environments and on ordinary surfaces. Special display configurations that require permanent projection canvases or artificial placeholders might become unnecessary.

To achieve this goal, however, several technical problems have to be solved:

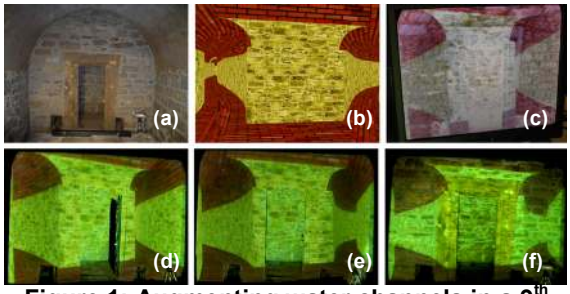

Figure 1: Augmenting water channels in a 9th century water reservoir (a): Screenshot of 3D model showing water pipelines behind wall (b), color compensated projection (c), registered augmented views allowing to see through wall (d,e), color uncompensated image (f).

1. View-dependent geometric correction for projecting images onto geometrically complex surfaces without causing image deformations;
2. Radiometric compensation for projecting images onto arbitrarily colored and textured surfaces to avoid color distortions;
3. Multi-focal projection for displaying images in focus on geometrically complex surfaces;
4. Seamless multi-projector blending for supporting immersive applications and enhancing image quality (e.g., brightness, color, elimination of shadows, etc.).

A correct and consistent depth perception of stereoscopically visualized images on everyday surfaces will fail in many situations if any of these problems cannot be solved adequately. Furthermore we believe that solutions to these problems have to be carried out on a pixel-individual level and in real-time to achieve practical results. In addition, calibration and registration efforts must be kept to a minimum.

In this paper, we want to discuss our first steps towards such an ad-hoc stereoscopic visualization in real environments. Inspired by unstructured Lumigraph rendering, we present an image-based warping method, as well as a refined geometric mapping technique for view-dependent projections onto geometrically and radiometrically complex surfaces. We outline our experiences made with radiometric compensation of stereoscopic projections onto colored and textured surfaces. Additionally, we describe first results of a multi-focal projection method that minimizes blur effects of images displayed on surfaces with varying depth. All these techniques are accomplished at

interactive rates and on a per-pixel basis for multiple interplaying projectors. Finally, we describe application examples and show how all these components can play together to support view-dependent stereoscopic AR/VR visualizations in real environments – without the need of special projection canvases.

We want to present these techniques in the remainder of this paper while discussing the related and previous work on individual topics within the corresponding sections. We do not describe seamless multi-projector blending techniques since these issues have been addressed earlier (e.g., [3, 4]).

2. Radiometric Compensation

If images are projected onto colored and textured surfaces, the projected light is blended with the reflecting surface pigments. This results in a fusion of projected image and underlying display surface texture.

Figure 1 illustrates this problem. Registered images of a three-dimensional scene are projected onto a natural stone wall. Figure 1f shows the blending of the uncorrected images with the underlying surfaces. A fusion of projected stereo-pairs –and consequently a correct depth perception of the displayed scene– is difficult in this situation. The reason for this is that the human visual system strongly relies on the extraction of salient structure features (such as edges, corners, etc.) for estimating disparities. Since for an uncorrected projection the features of the underlying surface can be dominant over features in the displayed images, the depth perception is not consistent. Portions with dominant real features (e.g., the visible real stones and mortar in figure 1f) lead to a perception of the real surface's depth in these areas. For areas in which image features are dominant, however, the depth of the virtual object can be perceived. Another example is shown in figures 3e-f. Thus a per-pixel radiometric compensation which neutralizes the blending artifacts is essential for visualizing stereoscopic imagery on textured surfaces (cf. figures 1d-e or figures 3g-h).

Recent work on radiometric compensation [3,10,16] uses cameras in combination with projectors for measuring the surface reflectance as well as the contribution of the environment light. These parameters are then used for correcting the projected images in such a way that blending artifacts with the underlying surface are minimized.

Nayar et al. [16], for instance, express the color transform between each camera and projector pixel as pixel-individual 3x3 color mixing matrices. These matrices are estimated from measured camera responses of multiple projected sample images. They can be continuously refined over a closed feedback loop and are used to correct each pixel during runtime. Later, a refined version of this technique was used for controlling the appearance of two- and three-dimensional objects, such as posters, boxes and spheres [10].

Another approach [3] uses single disjoint camera measurements of surface reflectance, environment light contribution and projector form-factor components for a per-pixel radiometric compensation using hardware accelerated pixel shaders. This method enhances radiometric compensation results by supporting multiple interplaying projectors. Figure 1c illustrates a corrected projector contribution and the observable results are shown in figures 1d-e.

A precise projection of pixels onto corresponding surface pigment is extremely important – especially for noisy surface texture, such as the stone wall in figure 1. Slight misregistrations can cause profound visual color and brightness artifacts, such as bright or dark spots that highlight the real surface features. The human visual system is extremely sensitive to such artifacts, and depth perception of virtual content is disturbed by highlighted real surface features. Note that geometric image distortions which result from such misregistrations are not as critical.

To achieve an adequate registration of projector pixels, we generate a 2D look-up table that maps every pixel from camera space to projector space and vice versa. We apply projected structured light patterns for measuring an unambiguous mapping. All radiometric measurements (surface reflectivity, environment light contribution and projectors' form factor components) are taken from the perspective of the camera. This ensures a direct look-up in the pixel shader for relating each projector pixel to the parameters of the corresponding surface pigment. Note that this principle is also being applied by related systems [10,16]. This mapping also allows a geometric image correction for the perspective of the camera if images are projected onto non-planar surfaces. In our approach, this is realized with a pixel displacement mapping in a pixel shader.

So far, we have discussed how images can be projected onto geometrically complex, colored and textured surfaces in such a way that they appear correct from the perspective of a single camera (or for observers located within the sweet spot area of the camera). These techniques can easily be extended towards multiple interplaying projectors to enhance the overall image quality [3], cancel out shadows cast by uneven surfaces [3] or carry stereoscopic images. For a single camera calibration, stereoscopic images can be rendered without having information about the viewpoint or the image plane using the method described by Jones et al. [12]. The stereo pairs can be projected with two or more projectors while stereo separation can simply be achieved with LC shutters in front of the projectors' lenses that are synchronized with active shutter-glasses worn by the users. However, stereo-capable DLP projectors provide high refresh rates (120Hz and more). They support active stereo with a single device and without a loss of brightness that is due to additional LC shutters. For view-dependent applications (e.g., head-tracked stereoscopic visualizations), however, a single sweet spot is not sufficient. Our extension towards a view-dependent correction of geometry and radiometric measurements is described below.

3. Geometric Warping

For projecting multiple undistorted images registered onto planar surfaces the estimation of projector-to-camera homographies is very common (e.g., [6]). Two main techniques exist for geometrically complex surfaces (a good overview can be found in Brown et al. [4]):

1. The per-pixel mapping technique described above is applied if multiple projectors have to be registered precisely for one fixed viewing position [3,10,16].
2. A 3D model of the surface can be acquired (either scanned or manually modeled). All projectors are then registered together with the surface model to a common coordinate system in which the observer is tracked. Projective texture mapping and two-pass rendering can be applied to geometrically warp the view of a moving user into the view of the projectors [19].

As we noted earlier, the first technique provides an adequate precision for a projection onto colored and textured surfaces but does not support arbitrary view points. The second method does support moving users, but suffers from registration errors and is not trivial to realize in practice [4]. Registration errors in the order of 2-3 pixels that lead to geometric distortions on white surfaces can sometimes be tolerated, in contrast to color and intensity artifacts caused on textured surfaces or inter-projector misregistrations. The geometry has to be scanned, and has to be precisely registered to its real counterpart. In addition the projectors have to be calibrated precisely to the surface. Both is difficult to achieve with the precision required by radiometric compensation. Finally, projective texture mapping is based on a simple pinhole camera model and can consequently not handle non-linear effects, such as radial distortion caused by projector lenses.

3.1. Image-Based Approach

To create a correct projection for a single camera, the geometric mapping between camera and projector(s), as well as the radiometric parameters have to be measured during calibration [3]:

Let *P2C* be a two dimensional look-up table in the resolution of the projector that maps every projector pixel to a corresponding camera sub-pixel (i.e., the image of the projected pixel on the surface). Each *P2C* map can be computed by reversing the mapping from camera space to projector space (*C2P*). A *C2P* map has the resolution of the camera and can be measured using structured light range scanning (gray code scanning with phase-shifting in our implementation). Note that *P2C* and *C2P* are usually not one-to-one mappings. Thus multiple entries have to be averaged during measuring and inversion – leading to a sub-pixel precise mapping. Furthermore when inverting a *C2P* map, the resulting *P2C* map might be incomplete. Thus missing portions have to be interpolated. This is done by triangulation and off-screen rendering into a 16 bit floating point P-buffer, followed by a read-back into the texture memory.

In addition to the geometric distortion, the radiometric parameters are measured. The surface reflectance and a projector's form-factor contribution are measured by projecting a white image and capturing the camera response (*FM*). The same is done for measuring the environment light as well as the black-level of each projector by projecting and capturing a black image (*EM*). Having an original image (*O*) and a definition of where this image has to appear within the camera view, a pixel shader can perform a pixel displacement mapping using the *P2C* map in such a way that *O* appears geometrically undistorted from the perspective of the camera. In addition, every pixel of *O* can be color corrected by looking up the corresponding radiometric measurements (*FM* and *EM*) and performing the necessary computations in the pixel shader. The parameter textures *P2C*, *FM*, and *EM* have to be measured for every projector that contributes to the final image. They are used by the pixel shader for computing the projector individual response. More details are provided in [3].

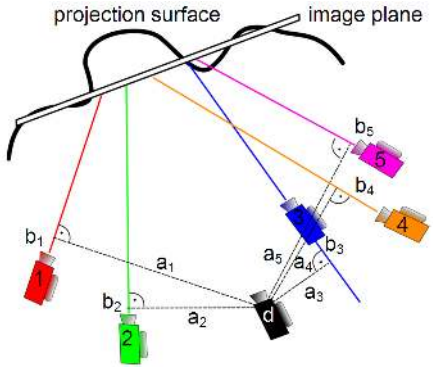

Figure 2: Image-based rendering approach: Five source cameras (1-5), one destination camera (d).

Figure 2 illustrates how this approach can be extended towards view-dependent image-based rendering that was inspired by unstructured Lumigraphs [5]:

We measure the set of parameters for multiple, unstructured source camera positions: P_i2C_j, FM_{ij}, and EM_{ij}, where i is the projector index and j is the camera index. Measuring one set takes about 20 seconds in our implementation. But this depends on the latency of the applied camera which –if no hardware synchronization is provided– has to be soft-synchronized to the structured light projection. The camera is tracked and we store its position and orientation together with each corresponding parameter set. Figure 2 illustrates this for five unstructured source camera positions. For rendering the image correctly, we have to define where the image plane will appear in space. This can be done once before calibration by interactively aligning a 3D model of the image plane at the desired position in the real

environment[1]. If the camera is moved, the registered image plane has to be rendered according to the new camera perspective. We offer two different image plane types: An on-axis image plane remains at a fixed position in space but its orientation is updated in such a way that it is perpendicular to the vector that is spanned by the camera's position and the central position of the image plane. The orientation and position of an off-axis image plane remains constant in space – no matter where the camera is located. The example in figures 3e-h show an off-axis image plane situation.

If we assume that we want to render a correct image only for one of the calibrated source cameras *j*, the following step is performed: The projection of the image plane into the camera's perspective has to be computed first. This is done by off-screen rendering the registered image plane model from the perspective of this camera. The image plane is shaded with texture coordinates that allow a correct perspective mapping of the original image *O* onto it. These texture coordinates range from u=0..1 and v=0..1 for addressing and displaying the entire image *O*. In the following we want to refer to this image plane texture as *IP*. Projector-individual pixel shaders can then carry out the following tasks: For each pixel of projector *i* find the corresponding radiometric parameters in FM_{ij} and EM_{ij} using P_i2C_j. Then find the corresponding pixel of the original image *O* by referencing P_i2C_j first to look-up the texture coordinate of *O* in *IP*. Using this texture coordinate, perform a look-up in *O*. Having all parameters, the color correction is performed [3] and the pixel is displayed.

For a novel destination camera position that does not match any of the source camera positions, however, all parameters have to be computed rather than being measured: The geometric and radiometric parameter textures, as well as the direction vector for this novel camera perspective are interpolated from the measured parameters of the source cameras. A new image plane texture *IP* is then rendered from this interpolated perspective. For a correct interpolation, the position of the destination camera is projected onto the direction vectors of the source cameras. Two distances can now be computed for each source camera *j* (cf. figure 2): The distance from the destination camera to its projection points on each source cameras' direction vector (a_j). And the distance from the destination camera's projection points to each source cameras' position (b_j). These distances are used for computing penalty weights for each source camera with:

$$p_j = \alpha a_j + (1-\alpha) b_j \qquad (1)$$

Note that all distances a_j and b_j have to be normalized over all source cameras before computing the penalties. The factor α allows weighting the contribution of each distance. Since a shift of the destination camera along a source direction vector causes less distortion than a shift away from it, a_j must be weighted higher than b_j. We chose $\alpha = 0.75$ in our implementation. Note that neither the orientation of the destination camera, nor the intrinsic parameters of source or destination cameras have to be taken into account for computing the penalty weights.

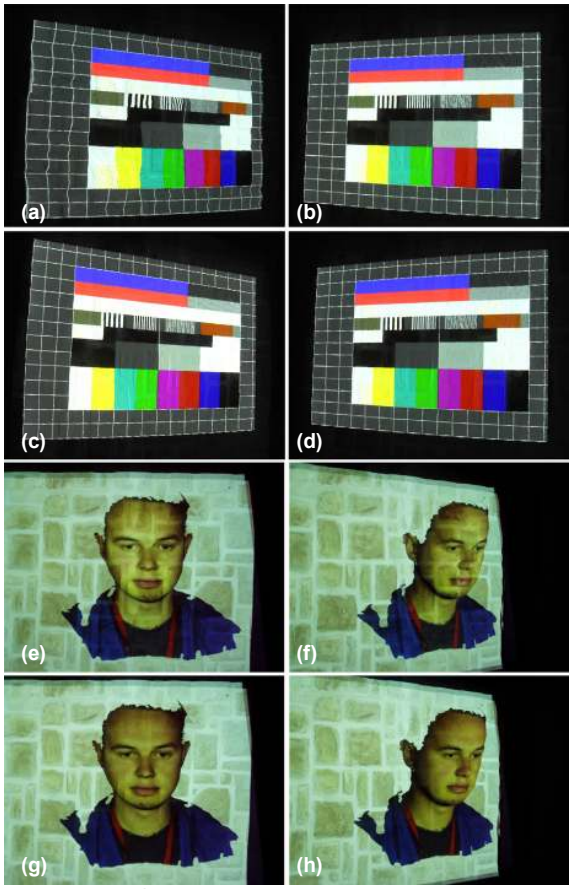

Figure 3: Geometric image warping: Image is distorted (a) when leaving the sweet spot of a single calibration camera (b), but not if a view-dependent correction is applied (c-d). Geometrically corrected stereo-pairs of 3D scene: without (e-f) and with (g-h) radiometric compensation.

From all source cameras, we select a subset of *k* cameras with the smallest penalties. Only these *k* source cameras are considered for sampling the destination camera's new parameter textures. For each of the *k* source cameras, a weight factor can be computed with:

$$w_j = \left(1 - \frac{p_j}{\max_{pk}}\right)\frac{1}{p_j} \qquad (2)$$

where $\max_{pk}$ is the maximum penalty among the *k* selected source cameras. Note that all weights have to be normalized after being computed. This implies that the source camera with the largest penalty ($\max_{pk}$) among the *k* selected ones is weighted with 0. A source camera with close-to-zero penalty is first weighted with a value

[1] Visual feedback for this process can be provided by rendering the image plane perspectively correct into the video stream of one source camera.

approaching infinity, but is then mapped to 1 after normalization.

The parameter textures and interpolated direction vector for the destination camera can now be computed, rather than being measured. This is performed with the pixel shaders by interpolating each parameter entry t_j of P_i2C_j, FM_{ij}, EM_{ij}, and the original direction vectors among the k selected source cameras as follows:

$$t_d = \sum_j^k w_j t_j \quad (3)$$

Note that look-ups in FM_{ij} and EM_{ij} have to be carried out with the original (non-interpolated) P_i2C_j map while look-ups in IP have to be done with the interpolated projector-to-camera map. This allows the computation of the geometric warping, the image plane projection, and the radiometric parameters (surface reflectance, environment contribution and black-level) for a novel destination camera. Note that for completely diffuse surfaces, the radiometric parameters do not change. Weak specular effects, however, are taken into account with this method as well. To handle a flexible number of source cameras, the pixel shaders are not hard-coded, but dynamically generated and loaded onto the graphics card during runtime. This happens only if k is modified.

Figures 3a-b show the image distortion that can be perceived if leaving the single sweet spot area (figure 3b). Figures 3c-d illustrate that these distortions are not present if the above described rendering method is used. Instead of applying an off-axis projection, simple texture mapping of the image on 2D quad on the cameras' image planes has been used in this example. Note that these images are all color corrected before projecting them onto the surface. Five source camera positions have been calibrated for this example.

Besides a static two dimensional image, interactively rendered stereo-pairs of a three-dimensional content can be displayed in O. These images can be rendered on-axis or off-axis for the defined image plane from the tracked position of the destination camera. This is shown in figures 3e-f for a color uncorrected scene and in figures 3g-h for the same scene with color correction enabled. Note that the tracked destination camera is equivalent to the tracked OpenGL camera defined for the perspective of the observer – not a physical camera. In the following we refer to it as the observer camera. Figures 1 and 8 illustrate other examples.

This image-based approach provides a pixel-precise mapping for view-dependent projections onto geometrically and radiometrically non-trivial surfaces. It does not require a 3D model of the surface or a registration of the projectors within a common three-dimensional coordinate system. However, it does correct non-linear projector distortions.

3.2. Geometry-Based Approach

If the geometry of the projection surface is known, a two-pass rendering technique can be applied for projecting the image undistorted [19]: In the first pass, the image that has to be displayed is off-screen rendered from the perspective of the observer (or the camera). This image O is then read back into the texture memory. In the second step, the geometry model of the display surface is texture-mapped with O while being rendered from the perspective of the projector. For computing the correct texture coordinates that ensure an undistorted view from the perspective of the observer projective texture mapping is applied. Using this hardware accelerated technique allows to dynamically compute a texture matrix that maps the 3D vertices of the surface model into the texture space of the observer's perspective.

It is easy to see that the precision of this method strongly depends on the quality of the surface model and on an adequate registration between surface model and projectors. As mentioned earlier, misregistrations of 2-3 pixels that lead to geometric image distortions can be tolerated. Performing the radiometric compensation with wrong parameters, however, leads to immediately visible color and intensity artifacts. They make the fusion of stereo pairs difficult. Below we describe a variation of this geometric rendering method that is applicable for view-dependent radiometric compensation:

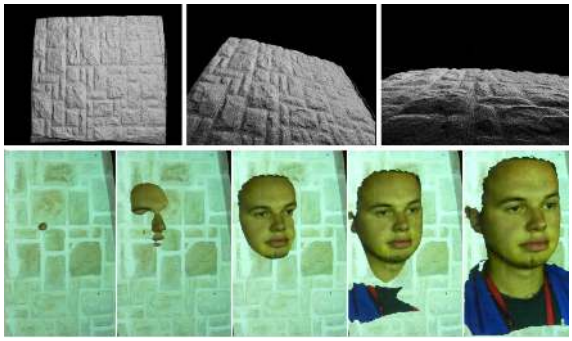

Figure 4: Computed geometry map (top) used for consistent occlusion between real and virtual objects (bottom).

For N camera perspectives, we compute N P_i2C_j maps for each projector. The P_i2C_j maps of any two camera positions lead to a single geometry map estimation that covers only the overlapping surface areas which can be seen from both cameras. A geometry map has the resolution of a projector. It stores 3D positions (in world coordinates) of the corresponding surface points (SP_i) at all projector pixels that project onto these points. The disparities for computing the surface points are given through a definite mapping from a single projector pixel to the two corresponding camera pixels. Consequently, $N^2/2$-$N/2$ partial geometry map estimations can be computed for each projector.

Using an unstructured and variable set of source camera positions, we find it hard to acquire a high-quality 3D model of the surface right away if the cameras are not mounted precisely. Small tracking errors (especially in orientation) lead to unusable initial geometry map estimations. We compute a final geometry map composition from all initial geometry map estimations by minimizing the geometric distance

of corresponding 3D points via a least-square optimization. The extrinsic camera orientations are therefore continuously re-adjusted until a global minimum is found.

Doing this for each projector allows us to compute viable projector individual geometry maps (GM_i) that cover all visible portions of the surface (cf. figure 4-top). They are aligned with their real counter part, but don't have to be explicitly registered to the projectors. A definite mapping of 3D surface points to projector pixels is provided implicitly through indexing GM_i. The radiometric parameters (EM_{ij} and FM_{ij}) are measured for each camera-projector combination. The parameter textures that belong to the same projector are then averaged and C_j2P_i-mapped to projector individual look-up textures (EM_i and FM_i) that correspond to the indexing of GM_i.

We compute a texture matrix (TM) that transforms the 3D surface points into the perspective of the observer camera. This matrix is a composition ($TM = N*I*E$) of extrinsic (E: position and orientation transformations) and intrinsic (I: perspective projection) parameters of the observer camera followed by a transformation from normalized device coordinates into normalized texture space (N=translate[0.5,0.5,0.0]*scale[0.5,0.5,1.0] for OpenGL). Note that the same matrix is also applied for texturing by conventional projective texture mapping methods. The rendering of the geometry from the perspective of the projector(s), however, is different in our approach:

During runtime a full-screen quad is rendered for each projector to trigger the fragment processing of every projector pixel. The following computations are then performed by the pixel-shader of each projector: For every projector pixel, the corresponding surface point SP_i is looked up in GM_i and is mapped into the perspective of the observer camera with $TM*SP_i$. As illustrated in figure 4-bottom, consistent occlusion effects can be achieved by performing a depth test with the transformed scene points and the depth map of the virtual scene. Being in the camera space, the pixel of the original image O can be referenced in the corresponding look-up textures, as described in section 3.1. Remember that for performing a look-up in O, the texture coordinates of the defined image plane have to be referenced. Thus the image plane texture IP has to be computed for the perspective of the observer camera and passed to the pixel shader exactly as described in section 3.1. The radiometric parameters can be looked-up in EM_i and FM_i. Note that EM_i, FM_i, and GM_i have projector resolution.

In contrast to conventional projective texture mapping approaches, this variation ensures that the look-up of the radiometric parameters for each projector pixel always matches with its corresponding surface pigment. Only the mapping into the observer's camera perspective depends on the quality of the estimated geometry map. This, however, can only cause a geometric misalignment of the image – but no color or intensity artifacts. Intrinsic and extrinsic parameters of the projectors do not have to be acquired. Non-linear projector distortions are corrected by this method as well.

4. Multi-Focal Projection

Conventional projectors can focus on single focal planes only. Projecting onto complex 3D surfaces with varying depth causes blur effects that wash out image features which are needed for correct disparity-based depth perception.

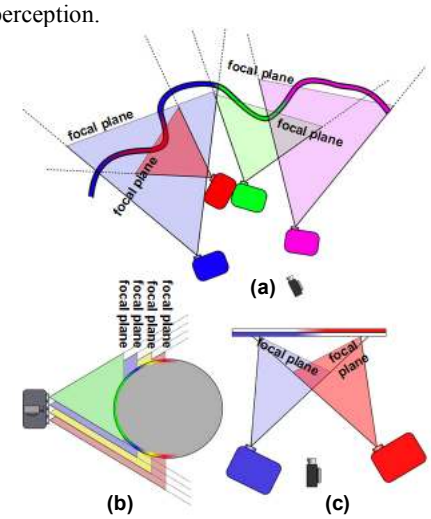

Figure 5: Multi-focal projection concept: Multiple projectors span different focal planes in space for different configurations (a,b,c). The focus value of each projector-pixel is estimated. A final image with minimal focus error is composed from multi-projector contributions.

To overcome this problem, planetariums or some VR displays [2] apply laser projectors instead of conventional projectors for displaying focused images on curved projection screens (e.g., domes or cylinders). Direct-writing-scanning-laser-beam projectors provide a very large focal depth. The cost of a single laser projector, however, can quickly exceed the cost of 500-700 conventional projectors.

Today's off-the-shelf projectors use structured light in combination with an integrated CCD sensor for measuring and adjusting an average focus of the entire image automatically to a planar projection screen. For projecting onto surfaces with varying depths, however, the focus of every single pixel has to be measured and adjusted individually.

Our multi-focal projection concept was inspired by the work on synthetic aperture confocal imaging of partially occluded environments [13], and by the work on creating graphical blur effects using superimposed projectors [15]. These approaches follow a reverse idea –creating synthetic defocus effects on real surfaces using projectors– and address different application goals. Akeley et al. [1], for instance, presented a desktop volumetric display prototype that provides multiple focal distances in space. Note that this is not possible with our solution, since we can only focus on the physical display

surface and not on arbitrary virtual planes in space that do not belong to the display surface. A focused stereoscopic projection on non-planar surfaces, however, allows a correct fusion of stereo pairs – but does not solve the general inconsistency between accommodation and convergence that is related to stereoscopic display approaches.

Our basic idea for achieving a multi-focal projection is illustrated in figure 5: Multiple conventional projectors with different focal planes, but overlapping images are used. They can be arbitrarily positioned in the environment (cf. figure 5a), or can be integrated into a single projection unit (cf. figure 5b). The focus error created on the surface has to be estimated for each projector pixel. If this is known, a final image with minimal defocus can be composed from the contribution of all projectors. We want to present our initial results below.

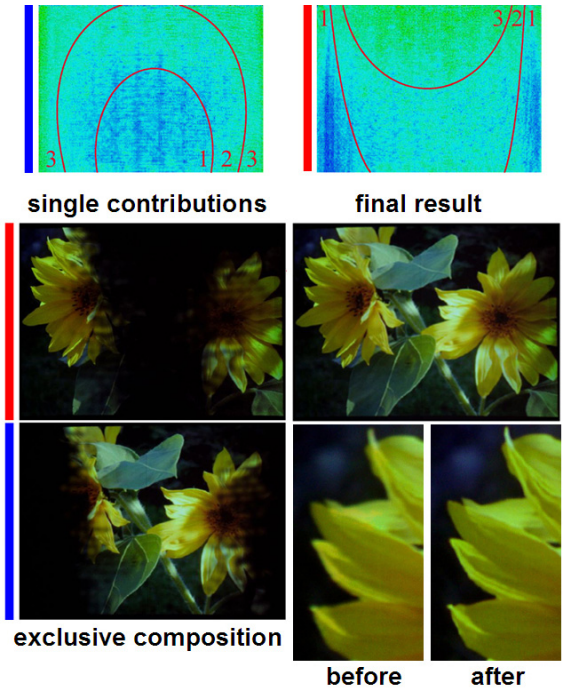

Figure 6: Initial multi-focal projection result: Two projectors (red and blue) project onto a cylindrical surface as illustrated in figure 5b. The blue focal plane intersects the surface in the front; the red focal plane intersects the surface in the middle. The measured focus values of both projectors (top) are used to determine the single pixel contributions that result in a final image with minimal defocus. Color coding: blue to green = best to worst.

4.1. Per-Pixel Focus Estimation

Several multi-focus image reconstruction techniques exist that compose a photograph with a virtually large focal depth from several registered photographs with differently focussed image portions (e.g., [9]). These algorithms have to estimate the focus values for each pixel using the image content only (e.g., via spatial image gradients). Estimating the focus values for an active display, however, allows simplifying this problem by analyzing displayed structured patterns (e.g., [21]). As mentioned above, such techniques are already being applied by off-the-shelf projectors for providing auto-focus capabilities, and are only capable of estimating an overall focus value for the entire image. Our intention, however, is to estimate an individual focus value for each pixel of each projector.

To determine the per-pixel focus values we project a uniform grid of sample points onto the projection surface in such a way that the points appear geometry and color corrected in the camera view. This is important because the focus estimation has to be independent of the projectors' and camera's distance to the surface, or the surface's color and texture. We achieve this by applying the geometry correction and radiometric compensation techniques described above. The result is a rectified and color corrected image of the projected sample point grid, whereby all points must have a uniform size, color and intensity – unless they are blurred because they are out of focus.

For estimating every sample point's blur we measure its color corrected and normalized point spread within a search window surrounding each point. The search windows have a uniform size and are also geometry corrected for the camera view. Based on the normalized point spread data, we have implemented and evaluated a variety of focus estimators that analyze the spread's area increase, and loss of intensities or high frequencies. The result is a relative focus value for every sample point. If the discrete sample points are then swept over the entire area of the search windows a focus value for every pixel in camera space can be estimated.

Figure 6 shows an example of two projectors with different focal planes creating an image of minimal defocus on a cylindrical surface (cf. figure 5b). The measured focus values (cf. figure 6-top) have been color coded and grouped into three areas (1=best, 2=medium, 3=worst focus) to give a better impression of the focus behavior. Note that the asymmetry along the vertical axis results from the vertical off-axis alignment of the projection frustum. Therefore, the image is always defocused more at the top than at the bottom. The circular shape of the focus areas are due to slight radial focus variations caused by the projector's lens system and due to the cylindrical shape of the surface.

The focus values are mapped into the perspective of the corresponding projector (via the *P2C* map) and are stored in a texture with projector resolution. We refer to this texture as focus map FOM_i. Note that the constant focal depth of the camera can increase the blur in our measurements. These camera specific focus effects are constant for all projector measurements. However, we want to compute only relative focus values that allow a comparison among multiple projector contributions rather than absolute values. Consequently, constant

camera-specific focus effects do not play a role for our algorithms.

4.2. Image Composition

Having the relative focus values for every projector pixel, a final image with minimal defocus can be composed and displayed from multiple projector contributions in real-time during rendering.

We offer two strategies for image composition: An exclusive (or binary) mapping ensures that a particular surface portion can be covered only by the projector that provides the smallest focus error at this portion. This method is shown in figure 6. The different contributions have to be blended adequately to take the different color responses of the projectors into account. The final composition image appears consistent and portions that were blurred before (i.e., with single projector contributions) are now in focus simultaneously.

Another strategy is to use the normalized focus values for weighting the pixel intensities of each projector. This allows overlapping projections and can consequently provide a higher image quality (e.g., brightness and contrast).

Note that both composition strategies are implemented as projector-individual pixel shaders, and the measured focus maps FOM_i are passed as parameter textures to the shaders. The focus values are view-independent. They are constant for each projector (if projectors and surface are fixed) and do not have to be estimated and interpolated for different perspectives, as it is the case for other surface parameters described in section 3.1.

5. Summary and Discussion

In this paper, we have shown how view-dependent stereoscopic projection can be enabled on ordinary surfaces (geometrically complex, colored and textured) within everyday environments. We pointed out several problems that have to be solved to ensure a consistent disparity-related depth perception and presented several solutions.

As illustrated in figure 7, we want to summarize that a consistent and view-dependent stereoscopic projection onto complex surfaces is enabled by four main components: geometric warping, radiometric compensation, multi-focal projection and multi-projector contributions. Implementing these components as hardware accelerated dynamic pixel shaders allows achieving interactive frame rates.

Radiometric compensation is essential for a consistent disparity-related depth perception of stereoscopic images projected onto colored and textured surfaces. We have outlined our experience made with stereoscopic projections of three-dimensional scenes using techniques that were previously used for correct projections of two-dimensional monoscopic images. Furthermore we have described how these techniques can be extended to support view-dependent rendering, and noted that an adequate registration precision is as important in this case as for fixed-view rendering.

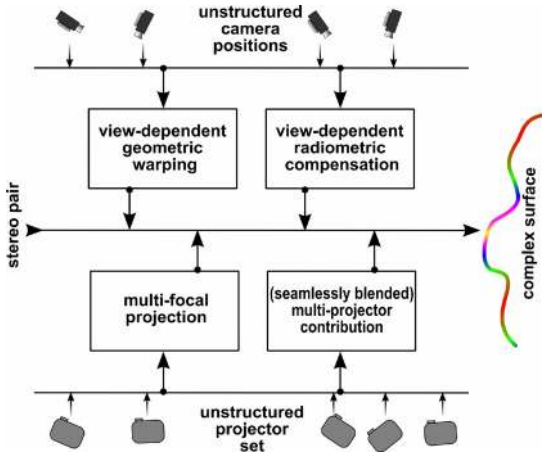

Figure 7: Main components for enabling consistent stereoscopic projection on everyday surfaces.

We presented an image-based warping technique and a refined geometric mapping method that both provide the required precision for radiometrically compensated projections of stereo pairs onto geometrically and radiometrically complex surfaces. Both techniques provide view-dependent rendering at interactive frame rates and support the correction of non-linear projector distortions.

The data that is required for the image-based method is relatively large (for k=N best camera positions: 3N textures for each projector plus 2 global textures), but is quick and easy to generate. Another advantage of this method is that geometric and radiometric parameters can be treated in exactly the same way. Such a method can serve as a general rendering framework that will allow computing additional surface parameters (such as view-dependent specular effects) in a uniform manner.

The complexity of this method, however, scales with the number of projectors and the number of camera positions. The data required for the geometry-based approach is relatively small (3 textures per projector plus 2 global textures) but is more difficult to produce. The reason for this is that the quality of the geometry maps depends strongly on the precision of the scanning technology. In case a tracked camera is used instead of a professional range scanner, the precision of the tracking system is a dominant factor. New radiometric parameters are not determined for novel observer positions. Thus view-dependent effects are not taken into account. The advantage of this method, however, is that its complexity scales only with the number of projectors. In addition, the geometry map allows creating consistent occlusion effects.

Note that the number of parameter texture (and consequently k) is limited by today's shader hardware. In general we found that for our experiments k=4-5 interpolation cameras (i.e., parameter sets) were sufficient to achieve good results with the image-based

method for large-scale projections (e.g., the examples shown in figures 3 and 8). This also approaches the maximum number of textures today's pixel shaders can process simultaneously. However, a manifold of source camera positions can be calibrated to cover a large projection space without affecting the rendering performance. Only the *k* best cameras are dynamically selected and passed to the pixel shaders for interpolation. For computing usable geometry maps, however, much more effort had to be put into. The rendering performances for both methods under the same conditions were comparable (32-16 fps for stereo with two projectors on an off-the-shelf PC with an nVIDIA FX6800 graphics card). A hybrid method that computes the image warping geometrically, but interpolates the radiometric parameters over discrete sample positions for taking view-dependent effects into account might be most efficient.

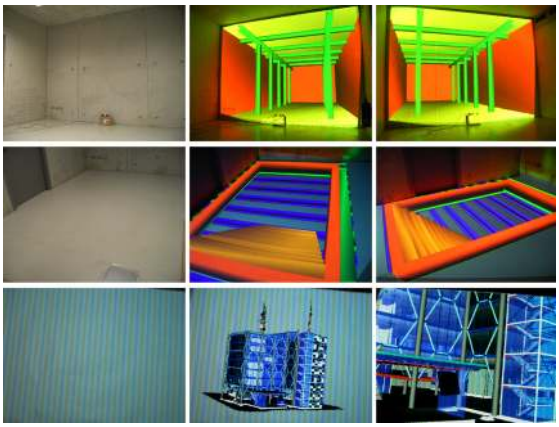

Figure 8: Application examples: Visualizing registered structures of an adjacent room (top) or stairways and lower building level (center) for architectural applications in a realistic environment. Displaying an architectural walk-through on a papered wall in an office (bottom).

We argued that a partially unfocused projection affects disparity-related depth perception and can make the correct fusion of stereo pairs difficult. Since conventional projectors provide only a single focal plane, it is physically impossible to generate sharp images on surfaces with extreme depth variations. We proposed a multi-focal projection method that provides a simple but efficient solution to this problem. Using multiple projectors, different physical focal planes can be generated. In this case, the entire problem can be solved in software – by automatically estimating per-pixel focus values and composing a final image with minimal focus error from multiple projector contributions in real-time. Since next-generation projectors will become very compact (size of a matchbox, using power efficient LEDs with low heat development instead of conventional light-bulbs) and cost effective, an integrated solution of many active stereo-capable DLP projection units into a single device (as outlined in figure 5b) is well imaginable. In this paper, we have described early results on our multi-focal projection concept. We are currently working on experiments with more complex surfaces using a denser set of focal planes.

To discuss multi-projector techniques (such as chrominance mapping, intensity matching, cross-fading, or shadow removal, etc.) was out of the scope of this paper. These techniques have been described in detail elsewhere. However, we do use these methods. In fact all described techniques are multi-projector capable and are used in multi-projector configurations.

As outlined in the architectural examples shown in figure 8 or the museum-oriented example shown in figure 1, such techniques do not only offer new possibilities for augmented reality and virtual reality. They also allow merging both technologies and give some applications the possibility to benefit from the conceptual overlap of AR and VR. An ad-hoc system configuration that is not constrained to laboratory conditions will increase the practicability of such a tool. The examples shown in figures 1 and 8 have been set up from scratch in a non-laboratory environment in less than one hour (including calibration and hardware installations). Note that the background of the visualized architectural scene in figure 1-bottom has been stenciled out and is intentionally not color corrected. This allows seeing through the 3D mode wherever possible to sense the real environment in addition. The effect is a consistent spatial disparity-based perception of the virtual and the real environment.

However, there are some technical limitations that have to be reported: The fixed resolution of both – cameras and projectors– prevent from measuring and correcting small geometric details and colored pigments that fall below their resolutions. The solution to this problem is to allow a higher spatial resolution (projector and camera) on a smaller surface area. This can be achieved simultaneously by a larger number of stationary projectors and cameras[2] or interactively by a single tracked projector/camera device. Furthermore off-the-shelf projectors and cameras suffer from a low dynamic range, which makes the capturing and compensation of a large color space impossible. Potential solutions are multi-channel projectors that provide a high-dynamic range in combination with multi-spectral imaging technology [11]. High-dynamic range or dynamic range increase techniques represent further software solutions that can enhance the camera measurements. The high black-level of conventional projectors also makes it difficult to produce absolutely dark areas. For multi-projector configurations the black-level of each projector is being added. This prevents us currently from using a large number of overlapping projectors for creating very bright images with a high contrast that can compensate all possible pigment colors. Optical filters can reduce the black-level – but they will also reduce the brightness. In some situations, local

[2] Or a single mobile camera covering a larger number of sample positions sequentially.

contrast effects (dark areas surrounded by bright areas are perceived darker than they actually are) reduce this problem on a perceptual level.

Diffuse materials that perfectly absorb light in one or more bands of the spectrum are not well suited for a radiometric compensation approach. Fortunately, such materials are not very common in the real world. Most diffuse surfaces scatter at least a small portion of the light being projected onto them. Thus this challenge reduces to the question of how much light we are able to project for achieving the desired result. Actually, it is the contrary problem whose search for solutions belongs to our future work: the sub-surface scattering of light that needs to be dynamically corrected for creating consistent immersive experiences in everyday environments.

References

[1] Akeley, K., Watt, S.J., Grishick, A.R., and Banks, M.S., A Stereo Prototype with Multiple Focal Distances, Proc. of ACM Siggraph'04, pp. 804-813, 2004.

[2] Biehling, W., Deter, C., Dube, S., Hill, B., Helling, S., Isakovic, K., Klose, S., and Schiewe, K., LaserCave – Some Building Blocks for immersive Screens, Proc. of Int. Status Conference on Virtual - and Augmented Reality, Leipzig, 2004.

[3] Bimber, O. Emmerling, A., and Klemmer, T., Embedded Entertainment with Smart Projectors, IEEE Computer, vol. 38, no. 1, pp. 56-63, 2005.

[4] Brown, M., Majumder, A., and Yang, R., Camera-Based Calibration Techniques for Seamless Multi-Projector Displays, IEEE Transactions on Visualization and Computer Graphics, vol. 11, no. 2, pp. 193-206, 2005.

[5] Buehler, C., Bosse, M., McMillan, L., Gortler, S.J., and Cohen, M.F., Unstructured Lumigraph Rendering, Proc. of ACM Siggraph'01, pp. 425-432, 2001.

[6] Chen, H., Sukthankar, R., and Wallace, G., Scalable alignment of large-format multi-projector displays using camera homography trees, Proc. of IEEE Visualization, pp. 339-346, 2002.

[7] Cotting, D., Naef, M., Gross, M., and Fuchs, H., Embedding Imperceptible Patterns into Projected Images for Simultaneous Acquisition and Display, Proc. of IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR'04), pp. 100-109, 2004.

[8] Ehnes, J., Hirota, K., and Hirose, M., Projected Augmentation – Augmented Reality using Rotatable Video Projectors, Proc. of IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR'04), pp. 26-35, 2004.

[9] Eltoukhy, H.A. and Kavusi, S., A Computationally Efficient Algorithm for Multi-Focus Image Reconstruction, Proc. of SPIE Electronic Imaging, pp. pp. 332-341, 2003.

[10] Grossberg, M.D., Peri, H., Nayar, S.K., and Bulhumeur, P., Making One Object Look Like Another: Controlling Appearance Using a Projector-Camera System, Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'04), vol. 1, pp. 452-459, 2004.

[11] Hill, B., (R)evolution of Color Imagining Systems, Proc. of 1st European Conference on Color in Graphics, Imagining and Vision (CGIV'02), pp. 473-479, 2002.

[12] Jones, G., Lee, D., Holliman, N., and Ezra, D., Controlling Perceived Depth in Stereoscopic Images, Proc. of SPIE, vol. 4297 Stereoscopic Displays and Virtual Reality Systems VIII, 2001.

[13] Levoy, M., Chen, B., Vaish, V., Horowitz, M., McDowall, I., and Bolas, M., Synthetic Aperture Confocal Imagining, Proc. of ACM Siggraph'04, pp. 825-834, 2004.

[14] Low, K-L., Welch, G., Lastra, A., and Fuchs, H.,Life-Sized Projector-Based Dioramas, Proc. Symp. Virtual Reality Software and Technology, IEEE CS Press, pp. 93-101, 2001.

[15] Majumder, A. and Welch, G., Computer Graphics Optique: Optical Superposition of Projected Computer Graphics. Proc. of Eurographics Workshop on Virtual Environment/ Immersive Projection Technology, pp. 209-218, 2001.

[16] Nayar, S.K., Peri, H., Grossberg, M.D., and Belhumeur, P.N., A Projection System with Radiometric Compensation for Screen Imperfections, Proc. of International Workshop on Projector-Camera Systems, 2003.

[17] Pinhanez, C., The Everywhere Displays Projector: A Device to Create Ubiquitous Graphical Interfaces, Proc. Ubiquitous Computing, Springer, pp. 315-331, 2001.

[18] Raskar, R., Welch, G., Cutts, M., Lake, A., Stesin, L., and Fuchs, H., The Office of the Future: A Unified Approach to Image-Based Modeling and Spatially Immersive Displays, Proc. of ACM Siggraph'98, pp. 179-188, 1998.

[19] Raskar, R., Brown, M.S., Yang, R., Chen, W., Welch, G., Towles, H., Seales, B., and Fuchs, H., Multi-projector displays using camera-based registration, Proc. of IEEE Visualization, pp. 161-168, 1999.

[20] Raskar, R., van Baar, J., Beardsley, P., Willwacher, T., Rao, S., and Forlines, C., iLamps: Geometrically Aware and Self-Configuring Projectors, Proc. ACM Siggraph, ACM Press, pp. 809-818, 2003.

[21] Tsai, D.M. and Chou, C.C., A fast measure for video display inspection, Machine Vision and Applications, vol. 14, pp. 192-196, 2003.