Orthographic Projection for Optical Signal Processing

Yves D. Jean

Orthographic Projection for Optical Signal Processing

Yves D. Jean

City University of New York (Lehman College)
yves.jean@lehman.cuny.edu

Abstract. Controlled illumination is a powerful tool for solving scene recognition problems. Binary and high frequency illumination primitives are projected into the scene to spatially encode or probe the optical environment.
Recently, computer vision researchers have shown that orthogonal functions and computational techniques from the signal processing framework can be mapped directly into the scene using projector-camera systems. These scene-space signal processing algorithms are achieved with illumination-encoded functions as primitives and computations derived from surface reflection models. Some examples of this new optical approach include convolution filtering and aliasing-canceling filterbanks.
In this paper we present a digital projector calibration method and an orthographic light field system. The calibrated system produces a light field structure matching the usual sampling geometry of image processing, well-suited to scene-space algorithms. The result is superior filtering and resolution performance due to higher accuracy optical representation. We evaluate our results by comparing processor-based filtering with a scene-space algorithm.

1 Introduction

Structured light techniques have proven to be flexible and powerful tools for scene analysis problems such as ranging and surface reconstruction (see Salvi et al [1] for an overview). They have been used to recover the geometric and optical properties of surfaces and to discern the global and local illumination components [2]. It is not surprising that controlled illumination has such capabilities because it introduces a known optical parameter into the irradiance that reaches the camera.

A common projector-camera technique is to encode the pixel positions of the projector with binary structured light to resolve point correspondences, via triangulation, with the camera pixels that recover the code [1]. Binary primitives, in general, minimize luminance transformations due to surface reflection, leaving a code mainly representing the spatial relationships between the projector, camera, and scene objects.

Techniques using high-frequency primitives have been used to probe optical properties such as defocus blur to compute depth [3,4]. A high frequency pattern

is projected into the scene and the reflected signal is analyzed for low-pass filtering effects (blur). There are other variations of these techniques highlighting the effectiveness of projector-camera systems in computer vision problems.

Recently, researchers have investigated light patterns based on orthogonal bases from the signal processing domain. The use of analytical functions as structured light primitives allows for new approaches to computer vision problems. The motivation is to perform computations directly in the scene by modeling signal processing formulations in the optical domain with digital projectors and cameras. The works of Damera-Venkata and Chang [5], Jean [6], and Ghosh et al [7] clearly demonstrate that signal processing formulations can be implemented in scene-space by optically encoding analytical functions and leveraging the native reflection models in the scene. This is the core computational model and is easily extended with additional optical components and image processing. Mapping signal processing tools into scene-space also requires satisfying the analytical constraints of sampling theory for signal analysis, representation and processing that must be preserved in the optical domain..

In this paper we present an orthographic projection calibration technique designed to produce higher accuracy light fields with only consumer-grade digital projector technology. The calibration technique corrects digital projector off-axis light field distortion and is applicable to optical systems incorporating digital projectors. The technique is especially effective for producing calibrated orthographic projection systems to bring the regular sampling geometry of signal processing to scene-space techniques. We present our results by implementing a scene-space algorithm and compare the results to processor-based image processing.

The paper is divided into the following sections. In Section 2 we review the related projector-camera literature and scene-space techniques, in particular. The calibration procedure is outlined in Section 3, and our results and analysis presented in Section 4.

2 Related work

There are many structured light techniques, using projectors and cameras, where a known illumination pattern is projected into the scene and the spatial or spectral transformation of the reflected signal is used to determine the configuration of objects and surfaces. We focus on coded light triangulation and depth from defocus, techniques built on a pattern primitive and related to our work with basis functions.

Salvi et al [1] provides an overview of optical triangulation coding strategy trade-offs depending on the computer vision performance objectives (i.e., segmentation ease, minimal image sequence, etc.). Binary coded light is easier to detect compared to other coding schemes though requiring more projections. Patterns built from binary primitives, like line stripes are commonly used in triangulation systems [8,9,10,11,12]. Smoother patterns such as ramps or gray level coding take fewer projections but increase the difficulty for code detection. However, sinusoidal primitives have been used with phase shifting to enhance the

resolution of recovered codes [13]. Binary coded primitives are flexible and have been used in other solution strategies. Young et al [12] provide an alternative to a sequence of pattern projections with multiple cameras in a view-point coding method. In Zhang et al [9] and Chen et al [14] multiple techniques are combined to minimize projection sequences and handle scene motion.

Depth from defocus is another approach to surface reconstruction using projectors and cameras. By measuring the extent of defocus on an imaged surface, the depth from the focal plane can be determined. Again, binary primitives assembled into a high frequency pattern are projected into the scene and reflected back to the camera along the same light path [4,3]. A coaxial configuration between the camera and projector enables the camera to directly image the high frequency pattern after traversing the depth dependent defocus volume in the scene. Another use of high frequency primitives is in an algorithm by Nayar et al [2] to separate the global and local radiance in a scene.

We present an orthographic projection system and calibration techniques. Other researchers have produced orthographic systems, for example, Nayar and Anand's [15] system drives a volumetric display with an orthographic light field. Lanman et al [11] use mirrors to transform an orthographic light field to produce an object surrounding structured light triangulation system. Neither [11] nor [15] address the projector distortion we examine in this paper, though the lens design optimization in [15] may reduce the effect. The calibration technique we present in this paper corrects the light field distortion in digital projectors. The technique is general and can be used in any system using digital projectors.

2.1 Scene-space techniques

Signal and image processing is a powerful analytical framework with a rich set of tools heavily used by the computer vision community to solve computational problems and for modeling the image formation process [16,3,4,17]. Researchers have developed novel new algorithms by shifting computations from the image to the scene directly. Scene-space techniques [5,6,7] implement signal processing computations in the optical domain by encoding scalar functions as illumination images, utilizing optical components, and modeling products and sums from the native surface reflection models or *bidirectional reflectance distribution function* (BRDF) in the scene. Equation 1 is the general surface reflection equation of the radiance B emitted from location $\mathbf{x}$ in direction ω_o due to the integration over the hemisphere parametrized by Z of ρ the BRDF and incident illumination L from direction ω_i (scaled by the surface orientation term $\cos\theta_i$):

$$B(\mathbf{x}, \omega_o) = \int_Z \rho(\omega_i, \omega_o) L(\mathbf{x}, \omega_i) \cos\theta_i d\omega_i \tag{1}$$

Damera-Venkata and Chang [5] showed that overlapped digital projectors can be used to produce super-resolution images on a Lambertian ($B(\mathbf{x}) = B(\mathbf{x}, \cdot)$) screen surface. Their approach takes advantage of the ability to produce aliased optical signals when the low-pass projector filter is insufficient [18]. They leverage the maximally decimated filterbank framework [19] to algorithmically determine

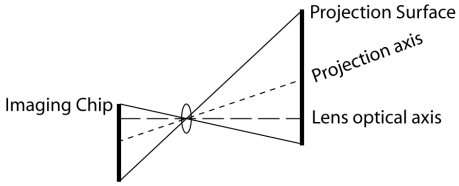


Fig. 1. Off-axis lens design of digital projectors [20]: The optical axis of the lens is above the center of the imaging chip resulting in a raised projection. The normally symmetric pyramidal projection is warped in the process, yet a correct image will appear on a projection surface parallel to the image chip.

the set of aliased images that, combined on a projection screen,

$$\hat{B}(\mathbf{x}) = \sum_i B_i(\mathbf{x}), \tag{2}$$

$(1 \leq i \leq \#projectors)$, cancel the aliasing errors.

In another example, Jean [6] presented a technique for optical convolution filtering by leveraging the intrinsic irradiance integration performed by camera pixels. The filter coefficients are embedded in the projector illumination and reflected from textured Lambertian surfaces. Lambertian reflection models a sample-wise multiplication of the surface texture $B(\mathbf{x})$ with the projected filter coefficients $f(\mathbf{x})$ which is integrated by a camera pixel,

$$\int_\Omega f(\mathbf{u})B(\mathbf{u})d\mathbf{u}, \tag{3}$$

(Ω = surface area) producing the filter response. The filtering technique computes the filter response in the scene before the lossy image formation process. A simple image capture records the camera pixel integration as the convolution integral. This advantage suggests it can be used to recover scene information within the extent of a pixel or super-resolution. Jean tested the technique on planar textured objects with Gaussian derivative filter images. The results show it can perform edge detection with no spatial processing of the images.

Ghosh et al [7] designed a BRDF acquisition system that performs a single operation: optical projection of the BRDF of a material sample onto spherical basis functions, with reflective optics and a projector-camera pair. It demonstrates the power gained with geometric warping of light fields to solve computer vision problems.

3 Orthographic projection

Scene-space techniques depend on digital projectors to convert image representations of scalar functions into a projection light field. The digital projector pixel grid defines the sampling geometry of the function and the distribution of rays

in the light field. However, the projection is problematic for two reasons: it is a pyramidal frustum, and thus, the light rays diverge in depth, and most digital projectors are designed to produce a distorted frustum for practical applications (Figure 1). In the follow subsection, we examine these problems and provide a calibration procedure for frustum distortion correction and an orthographic projector design.

3.1 Frustum distortion correction

Consumer digital projectors typically produce an off-axis projection frustum [20,21]. (Figure 1 shows a schematic of this optical design [20]) Off-axis optics produces a correct image on a wall or projection screen but distorts the light field as an artifact.

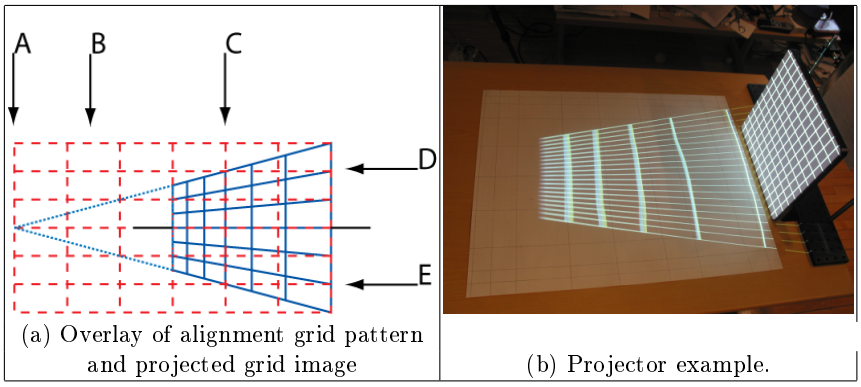
A simple solution to this problem is to use the lens shift feature found in some digital projectors. Lens shift is a mechanical control of the lens position along the vertical or horizontal direction. Figure 1 shows a vertical lens shift above the center of the imaging chip. Using this control, we can set the lens shift to zero (imaging chip and lens center aligned) creating an on-axis projection which produces a symmetric frustum. In Section 3.3 an optical design to transform a symmetric pyramidal projection to orthographic is provided. However, frustum distortion due to off-axis projection, if not removed with the following lens shift calibration process, will cause distortion in the orthographic light field.

3.2 Projector lens shift calibration

Each lens shift control (horizontal or vertical) independently translates the lens assembly along one axis of a plane parallel to the imaging chip. The independence between the two controls is key to our lens calibration procedure, the lens shift parameters are measured and corrected one at a time. Figure 2 illustrates and explains the approach we use to calibrate a digital projector with lens shift controls to find the zero lens shift position producing a symmetric frustum. The procedure requires a physical plane parallel to the lens shift axis to image the displacement. A flat table is used as the plane by simply placing the digital projector level on the surface. A regular grid pattern image is projected because it identifies groups of pixels to reveal frustum distortion. The regular grid pattern on the plane (dashed lines) intersects the projection image, imaging the frustum outline (solid lines). A solid line extending from the top and bottom of the frustum indicate the optical center. The dotted lines forming an apex are an annotated extrapolation from the truncated frustum of the projected image.

The objective of the calibration is to manipulate the lens shift control until the center line is aligned with a plane grid line, splitting the projected frustum into two symmetric halves. The results should mimic the symmetric frustum shown in Figure 2a, indicating that the axis lens shift is calibrated to zero displacement. Here is the step-by-step procedure:

Place a large grid patterned sheet on a flat table surface (or use a patterned surface or an optical bench). Set the adjustable feet of the digital projector such



(a) Overlay of alignment grid pattern and projected grid image

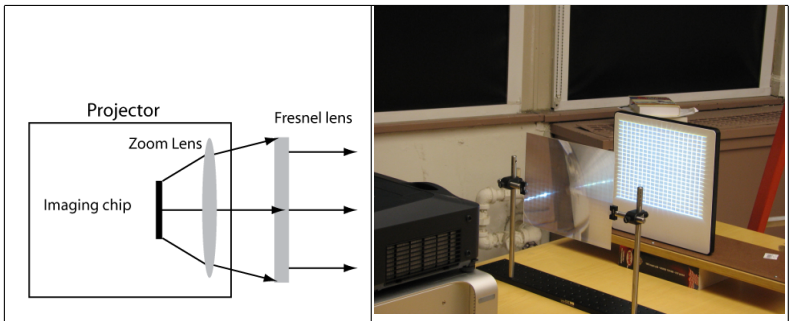
(b) Projector example.

Fig. 2. Lens shift calibration using a surface grid pattern and a grid image projection: (a) is an illustration of the projection frustum overlayed on the flat surface grid pattern (the dashed lines). The projection image, shown as solid lines, is a grid pattern that intersects the surface revealing a truncated frustum used to establish frustum symmetry. Five landmarks in (a) represent the design parameters for an orthographic projection device (see Figure 3). Landmark A is the center of projection. B and C represent the front of the digital projector (zoom lens housing) and the Fresnel lens mount position, respectively. D and E delimit the lens aperture. (b) shows our projector calibration setup (projector out of view). The vertical object on the right, is a Spectralon target.

that the base sits flat. Position the patterned sheet to intersect the light output from the digital projector producing a (initially distorted) frustum. Figure 2b shows an example. Here is the step-by-step

1. Choose the lens shift control which manipulates the displacement axis parallel to the table surface (when the projector is sitting on the foot pegs this will be the horizontal lens shift controls).
2. Align the center line of the projected grid image with a line on the loose patterned sheet. (see Figure 2b)
3. Simultaneously manipulate the lens shift control and adjust the patterned sheet until the projected grid image produces a symmetric frustum about a line on the surface grid pattern (see Figure 2a). The axis is now calibrated.
4. Turn the projector onto the side and repeat the procedure beginning with Step 1 to calibrate the other axis.

Perform the following to verify that the optical axis is parallel to the table surface. Measure the height of the vertical center of the light frustum exiting the digital projector lens. Using a vertical surface as a reference (see Figure 2b, we use a Spectralon target), check that the height of the center horizontal grid image line is equal to the previous measurement and level with the surface. Note that camera calibration [22] of the projector can be performed to measure radial lens distortion. We did not find it necessary if lens shift calibration using the center grid image lines is performed properly.



a) Schematic b) Orthographic light field test

Fig. 3. Orthographic projection device: a) is a schematic of an orthographic projection device, a projector and Fresnel lens form an optical combination to produce an orthographic field. Figure (b) shows a completed assembly. A grid image is projected onto the Fresnel lens and appears unscaled a distance away on a reflective surface, verify an orthographic light field.

The orthographic projection systems, [11,15], both use fixed off-axis projectors but do not address frustum distortion correction or compensation. Frustum distortion was not a factor in Damera-Venkata and Chang's work because the scene-space computation occur on the projection screen. Ghosh et al did not address nor correct for frustum distortion with their reflective design. However, reflective optics is an alternative [20] to the refractive (Fresnel lens) optics approach presented in the next section. Jean's algorithm is affected by frustum distortion and we will show how calibration improves the resolution and filtering performance.

3.3 Projection transformation optics

The symmetric pyramidal light field from an on-axis digital projector can be transformed into an orthographic light field with refractive optics (see Figure 3). A Fresnel lens is a collimator of a point source or, equivalently, the rays from a pyramidal light field. Nayar and Anand [15] and Lanman et al [11] also used Fresnel lenses in their orthographic systems. Figure 3 shows a schematic for the orthographic system design and a picture of our setup. We now provide a calibration procedure to determine the focal length and lens positioning. The initial step is to establish orthogonal spatial relations amongst the components. This calibration procedure is a continuation of Section 3.2 and is similar to steps outlined in [11].

To produce an orthographic light field with a digital projector we must establish two orthogonal planes to position and place the projector and other components. A third orthogonal plane will define the optical axis and fix the orientation of the components. A flat surface, upon which the projector is placed, defines one plane. Using a calibrated frustum from the previous section creates

a parallel relationship between the digital projector's optical axis and the surface. A vertical planar object (a Spectralon target) defines the second plane. The Spectralon target and the flat surface form an orthogonal pair of planes, however, the light field frustum from the digital projector must be aligned and integrated into this orthogonal relationship.

The lines of the grid image on the flat surface between the digital projector and the target (see Figure 2b) are used to align the two objects. With the projector in a fixed position, the center horizontal and vertical grid image lines are orthogonal and define two planes that intersect about the optical axis. The center vertical line is used to position the Spectralon target, relative to the optical axis, and the horizontal lines along the axis are used for perpendicular orientation. This alignment method is used in the following orthographic system design process.

Figure 2a shows landmarks A, B, C, and D, relative to the frustum of a digital projector. The figure illustrates and defines the five landmarks which are the parameters necessary to construct an orthographic projection device as shown in Figure 3. Given A, B, C, and D we can calculate the focal length ($focal\ length = |A - C|$), aperture ($aperture = |D - E|$), and mounting position ($lens\ mount\ displacement = |B - C|$) of the Fresnel lens. Replacing the Spectralon target with a Fresnel lens (see Figure 3b) converts the projector's light field to orthographic. The vertical and horizontal center lines of the grid image must align about the center of the Fresnel lens.

The choice of Fresnel lens parameters and mount position are determined by varying the digital projector zoom control to find the largest image size and position matching a commercially available Fresnel lens aperture and focal length. Once the design parameters are determined, a system like Figure 3b can be assembled. We used an 8.5" X 10.5" aperture Fresnel lens with an 18" focal length (Edmund Optics #43-015) in the figure.

4 Results

We tested the orthographic projection system by implementing Jean's [6] scene-space filtering algorithm to perform edge detection. We used a ViewSonic PJ1158 3LCD digital projector with dual lens shift controls, and an IMPERX VGA-210 monochrome 1CCD camera (12-bit resolution). The test images and filter response images are shown in Figures 4 and 5. The test image in Figure 4a, is printed and mounted on a flat panel and placed in front of the orthographic projection system (the projector focus plane was set at the print surface). Figure 4b is the camera image of the print from the 1CCD camera, this camera will be used to perform filter integration as defined in Jean's algorithm. Figure 4c is taken with a higher resolution 3CCD camera with a zoom lens to highlight the projected filter pattern composed of a grid of Gaussian derivative primitives (images) [6].

First we compare filtering performed on a computer processor to orthographic scene-space filtering. A pair of orthogonal Gaussian derivative edge detection filters are used in both cases to compute edge energy [23]. Using Matlab, the

(a) Original image (b) Camera image (c) 3CCD closeup view of filter projection

(d) Monochrome (e) Computer filtering (f) Camera image (g) Optical filtering

Fig. 4. Test of the orthographic system: (a) is the original test image. (b) is the image mounted on a poster board and viewed through a monochrome camera. (c) is a close up of Gaussian derivative filters ($7 \times 7, \sigma_x = 2, \sigma_y = 2$) projected on the poster and viewed through a 3CCD camera. (d) is a monochrome version of the test image. (e), computed using Matlab, is a threshold image after filtering (d) with the Gaussian derivative filters. After printing and mounting (d) on a board, it is illuminated with uniform illumination (box functions) and viewed through a monochrome camera (e). (f) is a scene-space filter image with our orthographic projection system and using Jean's algorithm and the same Gaussian derivative filters.

(a) Camera image (b) Edge threshold (c) Threshold #2 (d) Blue channel edges

(e) Camera image (f) Edge threshold

Fig. 5. Further examples of orthographic scene-space edge detection (Note: Figures 4b and (a) are the same): (a) represents the test image (Figure 4f) at another scale, and revealing new features. (b) is an arbitrary threshold of the energy image. (c) is another threshold level of the energy image. (d) is an image of the test image under an only blue color spectrum filter projection. Notice the yellow siding of the building in the background disappears because the blue illumination is not reflected.

test image is converted to monochrome (Figure 4d) and convolved with the derivative filters. An arbitrary energy threshold was selected to produce the computer filtering reference image shown in Figure 4e. A mounted print of the test image is viewed by the monochrome camera in Figure 4f. After applying the Gaussian derivative scene-space filters an arbitrary threshold is chosen to produce the image in Figure 4g. Using the edges in Figure 4e as a baseline for comparison, note that the scene-space filtered edges in Figure 4g are similar. The derivative filters perform as expected: constant areas produce zero filter response and edges are enhanced. Edge details are found in the face, hair, hand, as well as the background. This result shows that the regular sampling geometry of orthographic light fields can produce computational results comparable to sampling with a computer processor.

The test image is printed at another scale in Figure 5a to show edge performance under scale change and to expose other parts of the image. Two energy thresholds were chosen, Figures 5b and 5c, to demonstrate the range of available

edge information. The two edge images illustrates the resolution available with orthographic light fields under fine to coarse scene sampling. Please note that Jean used a Bayer mosaic camera in their paper, but in our implementation we use a monochrome camera to demonstrate how spectral control can be used in scene-space filtering. Figure 5d shows the result of applying blue color-only coded derivative filters to Figure 5a. The edge image differs from the previous two results because the filter response is limited to the blue spectrum of the BRDF. Note the yellow (red+green) siding of the house in the background produces no filter response except at the seams where the color mix shifts to gray. Another test image is shown in Figures 5e and 5f, containing a zebra pattern and a natural background. Naturally, the edge image shows the zebra pattern is enhanced but also shows the background features are segmented as well.

The scene-space filtering tests are not meant to represent computational accuracy. Neither the projector illumination nor the camera response were subject to photometric calibration. The BRDF of the prints were not analyzed but approximated as a diffuse material embedded with an albedo modeled ink. The goal was to show that even under these sub-optimal experimental conditions strong filter responses were measured.

There are optical performance limitations in our orthographic system design. For example, Fresnel lenses introduce optical artifacts and are not well-suited for imaging tasks. Digital projectors have large apertures that limit the depth of field and orthographic projection transformation does not remove this problem. Thus, there is a practical limit to the effective size of the working volume of the orthographic projector light field. We expect the emerging solid-state RGB laser projector technology will solve the depth of field problem. Also, our results are based on a planar test object, generalizing to curved objects with large surface undulations is a natural next step. We will attempt to handle these conditions by the use space-coding techniques to determine surface geometry before projecting filters. This approach can also be used to align camera pixels with filter primitives.

5 Conclusion

We have presented an orthographic light field system design and calibration method using consumer digital projector technology. A calibrated orthographic light field has an optical sampling geometry consistent with signal processing theory. The efficacy of our methods were tested by implementing a scene-space algorithm and clearly show results comparable to computer processor-based computations. The depth of field limitation must be solved before our method can fully take advantage of signal processing theory and algorithms. Bridging the two domains will enable new computer vision techniques that leverage signal processing algorithms directly in the scene. Our next step is to increase the accuracy of optical computation techniques by compensating for surface orientation.

References

1. Salvi, J., Pages, J., Batlle, J.: Pattern codification strategies in structured light systems. Pattern Recognition **37** (2004) 827–849
2. Nayar, S.K., Krishnan, G., Grossberg, M.D., Raskar, R.: Fast separation of direct and global components of a scene using high frequency illumination. ACM Trans. Graph. **25** (2006) 935–944
3. Nayar, S.K., Watanabe, M., Noguchi, M.: Real-time focus range sensor. IEEE Trans. Pattern Anal. Mach. Intell. **18** (1996) 1186–1198
4. Zhang, L., Nayar, S.: Projection defocus analysis for scene capture and image display. In: SIGGRAPH, New York, NY, USA, ACM (2006) 907–915
5. Damera-Venkata, N., Chang, N.L.: Realizing super-resolution with superimposed projection. In: CVPR. (2007)
6. Jean, Y.: Scene-space feature detectors. In: CVPR: Beyond Multivew Geometry. (2007) 1–8
7. Ghosh, A., Achutha, S., Heidrich, W., O'Toole, M.: Brdf acquisition with basis illumination. In: IEEE International Conference on Computer Vision. (2007)
8. Scharstein, D., Szeliski, R.: High-accuracy stereo depth maps using structured light. (2003) I: 195–202
9. Zhang, L., Snavely, N., Curless, B., Seitz, S.M.: Spacetime faces: high resolution capture for modeling and animation. In: SIGGRAPH, New York, NY, USA, ACM (2004) 548–558
10. Zhang, S., Huang, P.: High-resolution, real-time 3-d shape acquisition. In: IEEE CVPR Workshop. Volume 3. (2004) 28–37
11. Lanman, D., Crispell, D., Taubin, G.: Surround structured lighting for full object scanning. In: 3DIM07. (2007) 107–116
12. Young, M., Beeson, E., Davis, J., Rusinkiewicz, S., Ramamoorthi, R.: Viewpoint-coded structured light. (2007) 1–8
13. Huang, P.S., Zhang, C., Chiang, F.P.: High-speed 3-d shape measurement based on digital fringe projection. Optical Engineering **42** (2003) 163–168
14. Chen, S., Li, Y., Zhang, J.: Vision processing for realtime 3-d data acquisition based on coded structured light. **17** (2008) 167–176
15. Nayar, S., Anand, V.: 3d display using passive optical scatterers. IEEE Computer **40** (2007) 54–63
16. Durand, F., Holzschuch, N., Soler, C., Chan, E., Sillion, F.X.: A frequency analysis of light transport. In: SIGGRAPH, New York, NY, USA, ACM (2005) 1115–1126
17. Ramamoorthi, R., Hanrahan, P.: A signal-processing framework for reflection. ACM Trans. Graph. **23** (2004) 1004–1042
18. Damera-Venkata, N., Chang, N.L.: On the resolution limits of superimposed projection. Image Processing, 2007. ICIP 2007. IEEE International Conference on **5** (Sept. 16 2007-Oct. 19 2007) V –373–V –376
19. Mallat, S.: A Wavelet Tour of Signal Processing. Academic Press (1998)
20. Smith, W.J. In:. Third edn. McGraw-Hill (2000)
21. Raskar, R., Beardsley, P.A.: A self-correcting projector. In: CVPR (2), IEEE Computer Society (2001) 504–508
22. Bouget, J.: Camera calibration for matlab. Technical report, http://www.vision.caltech.edu/bouguetj/calib_doc/ (2001)
23. Perona, P., Malik, J.: Detecting and localizing edges composed of steps, peaks and roofs. (1990) 52–57