3D-Color Video Camera

O. Rubinstein Y. Honen A. M. Bronstein M. M. Bronstein R. Kimmel

Geometric Image Processing Laboratory
Computer Science Department
Technion -IIT, Haifa 32000, Israel

soffer|yaronh|bron|mbron|ron@cs.technion.ac.il

Abstract

We introduce a design of a coded light-based 3D color video camera optimized for build up cost as well as accuracy in depth reconstruction and acquisition speed. The components of the system include a monochromatic camera and an off-the-shelf LED projector synchronized by a miniature circuit. The projected patterns are captured and processed at a rate of 200 *fps and allow for real-time reconstruction of both depth and color at video rates. The reconstruction and display are performed at around* 30 *depth profiles and color texture per second using a graphics processing unit (GPU).*

1. Introduction

Accurate geometry acquisition of moving objects is required in a wide variety of applications ranging from marker-less motion capture to three-dimensional face recognition. Despite the availability of numerous acquisition and reconstruction methods, achieving high spatial resolution at video rates with a low cost system is still a challenging problem. Many traditional acquisition techniques like laser scanning [4] that produce very good results in static scenes are too slow and, therefore, unsuitable for rapidly changing scenes. Among techniques allowing fast acquisition a notable place is taken by the *time of light* approach introduced recently by 3DV-systems [8]. This method measures the depth profile of reflected photons resulting from an accurately gated and synchronized light pulse illuminating the scene. Time of light has low complexity reconstruction and can be implemented by a fast camera with linear response achieving video rates. However, as it involves compensating for the object albedo and other optical and manufacturing factors, its spatial resolution has so far been limited to a few centimeters.

Other techniques allowing fast acquisition are the *shape-from-stereo* methods based on multiple views. However, the simplest *passive stereo*, unaided by external sources of illumination, involves the non-trivial challenge of finding dense correspondences between two images of the same scene captured at two different viewing angles. Accurate solution of the correspondence problem might be prohibitively computationally intensive, and even then the obtained spatial resolution is usually mediocre. For that reason, passive stereo techniques can be found in a very limited set of applications.

A large family of *active stereo* techniques use controlled source of illumination, which allows to alleviate or completely resolve the correspondence problem. This makes active stereo attractive for accurate high-speed acquisition. We found that a good tradeoff between acquisition and reconstruction speed, reconstruction accuracy, and build of materials cost is achieved by the *coded light* approach, a popular active stereo technique in which one of the cameras is substituted by a controlled projector. The projector illuminates the scene with a pattern allowing to decode the projection plane at every point. This solves the correspondence between the camera and the projector, thus reducing the reconstruction effort to simple two-view triangulation. This technique is both robust and accurate when the pattern is time-multiplexed (i.e., a sequence of patterns is used). Yet, as more than a single pattern is required for fine coding of space, the reconstruction of a depth profile at video rates can be challenging.

In this paper, we revisit the ingredients of coded light acquisition, and present a design of a cheap and accurate 3D camera operating at video rates.

2. Existing technology

A recent review on coded light techniques can be found in [14] and a detailed description on pattern coding and reconstruction algorithms in [5]. In time-multiplexed coded light systems, the light code is constructed by a sequence of patterns projected on the object. Therefore, the main challenge in such systems is encode spatial location in minimum projection time.

Several solutions targeting low-cost accurate depth acquisition based on coded light have been described in the literature. In an early effort by Rocchini et al. [10], color is acquired separately, no real-time frame rates are reported, and the reported accuracy is one order of magnitude less than the one we report here. In a more recent work by Vieira et al. [13], the scanning system is based on coding part of the projected patterns in color. This approach requires a color camera and can be sensitive to the reflectance properties of the scanned objects. Moreover, accurate reconstruction occurs at the transaction between color stripes, rather than a specific code for each projected pixel. Both properties, a color camera with Bayer filter and chromatic multiplexing, impose limitations on the potential resolution of such a scanning system.

In our design, we use a time-multiplexing DLP projector, in which color image is constructed by high frequency sequential projection of the three color planes. This allows us to compress three different code planes into a single color image projected in one refresh cycle. In order to further compress the number of patterns required to encode the spatial location, we manipulate the frequency of projection of each pattern, allowing for fast acquisition of fine structures containing high spatial frequencies, while performing slower acquisition of low spatial frequencies. A monochrome camera is used to acquire the patterns. Since each pixel is sequentially illuminated by the three base colors, we are able to reconstruct color texture of the scene.

3. System components

Optimizing for cost per pixel per second, we chose a monochrome PointGrey Dragonfly Express camera [1, 2, 3], configured to approximately 180 fps at a resolution of 360×480. The camera maximal resolution is 480×640, but configured at frame rates above 120 fps and using an external synchronization signal, it transmits only 70% of the image without compromising for the other parameters. For generating the coded light we use a Toshiba DLP LED-based projector with native resolution of 800×600. The camera is triggered through a signal obtained by manipulating the projector's native 60Hz clock. Three output patterns that correspond to red, green, and blue are projected at each refresh cycle. A software application controls pattern projection, image grabbing, processing, reconstruction and display, and produce continuous real-time 3D color video stream. Processing is performed on a Intel PC with four dual-core 2.40GHz CPUs, 2.5GB RAM and an NVIDIA GeForce 9800 GX2 GPU with 0.5GB memory. The acquisition device with a 6 inch ruler are depicted in Fig. 1.

Figure 1. Our desktop 3D video camera

4. Fast 3D reconstruction

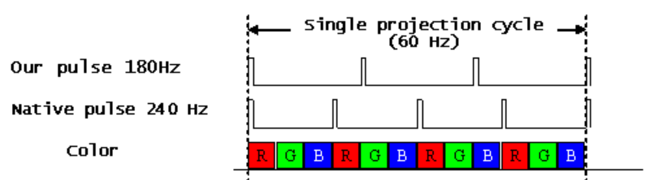

Figure 2. Synchronizing projector pulse and camera grabbing.

The DLP projector refresh rate is 60Hz, where each projection cycle is composed of four 240Hz RGB sub-cycles. That is, color planes are interleaved and each of them is projected four times for approximately 1.3ms. We use a custom circuit that divides the projector's inner 240Hz clock signal into a synchronized 180Hz signal, which triggers the camera and allows grabbing the distinct color planes of the projected image at constant time intervals (Fig. 2).

At startup, the initial location of the trigger within the projected sequence of patterns captured by the camera is unknown. This causes some uncertainty in the order of capturing the color planes. The captured sequence can therefore be RGB, GBR or BRG as shown in Fig. 3. A pattern detector is therefore applied in order to identify the shift and synchronize at the exact order of the color planes. Matching the acquired image to its corresponding colored pattern is necessary for proper reconstruction of the projected code. For that goal, we use a phase locking synchronization loop, in which among all possible reconstruction combinations the one with the highest local monotonicity of the stripe code is selected.

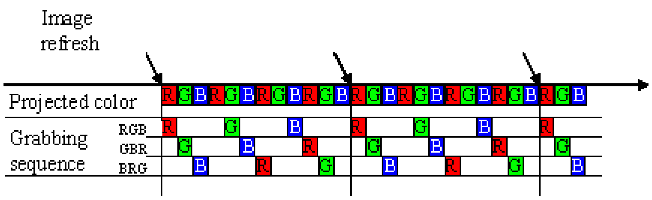

Figure 3. Synchronizing with the projected patterns.

4.1. Projected patterns

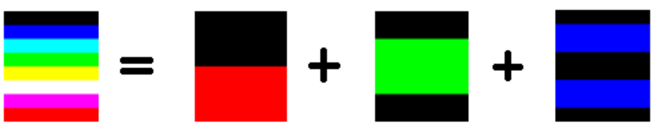

Figure 4. Projected image (left) as a composition of three distinct stripe codes.

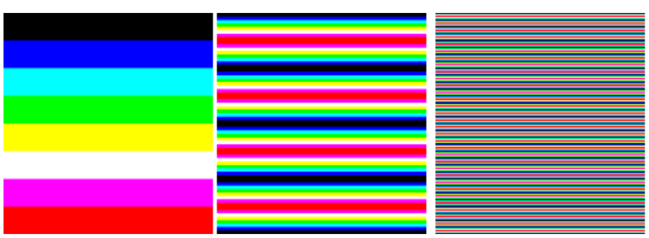

Figure 5. Three projected images disguising nine patterns.

We use the acquisition sequence described above with a sequence of 18 patterns composed of nine Gray code patterns (Fig.s 4 and 5) and their corresponding negative images[1]. Each sub-sequence is sufficient to construct a 3D image. Using this setup, 20 color frames per second are acquired at 0.1mm spatial resolution for stationary objects.

In order to further accelerate the frame rate, it is desirable to reduce the number of patterns while preserving the reconstruction accuracy. We accelerate the rate from 20 depth fps to 30 depth fps using the same hardware as follows: Each new sequence contains 24 patterns from which we reconstruct four 3D images (Fig. 6). The 24 patterns are composed of five high frequency patterns (least significant Gray code bits) and four low frequency patterns (most significant Gray code bits). Each one of the four low frequency patterns appears only once every 24 patterns, while each of the five high frequency patterns appears once every six patterns. As before, each projection is a composition of three distinct patterns. This way, we reduce the number of projections, thus enabling an accurate 3D reconstruction every two projection cycles.

[1] This idea that is used to extract the running threshold level for each color channel was suggested by Dr. Doron Shaked of HP Labs.

4.2. Reconstruction

Once a complete sequence of images is collected from the camera, decoding the Gray code and subsequent triangulation are performed by an HLSH DirectX shader. First, for each pixel in the 360×480 image grabbed by the camera, the stripe code is extracted. The shader function computes the per-pixel maximum and the minimum values of all patterns of the same color, which are used as local binarization thresholds. Binarization yields the bits of the Gray code at every pixel. Next, the known camera and the decoded projector coordinates are passed to the reconstruction shader function, which performs triangulation using the camera and projector matrices, pre-computed at the calibration stage [5]. The output of the reconstruction stage are the spatial coordinates of each pixel.

The shader responsible for decoding and reconstruction contains 289 instructions (5 of which are memory I/O and the remaining 284 are arithmetic operations). Sub-pixel resolution can be achieved by interpolating between the coded stripes. When used, the processing is split into two shaders due to instruction count limitations.

Due to camera noise and imperfections along stripe-code boundaries, filtering the 3D reconstructed profile is sometimes required. Among different robust filters, the best results are achieved by the Beltrami filter [11, 6] that takes into consideration the gradient of the reconstructed profile, or an efficient 2D median filter.

4.3. Calibration

For a review of camera calibration methods the reader is referred to [7, 12, 15]. We used Zhang's two camera calibration approach with a some adjustments [15]. Following Zhang and Huang [14], we model the projector as a second inverted camera.

$10-14$ images of a checkerboard at different locations and orientations are captured; $5-7$ are used for the calibration and the rest for verification. A pinhole camera model is assumed as well as standard calibration parametrization as described in [9]. Optical distortion is estimated and corrected for at the initialization phase. Using the checker-

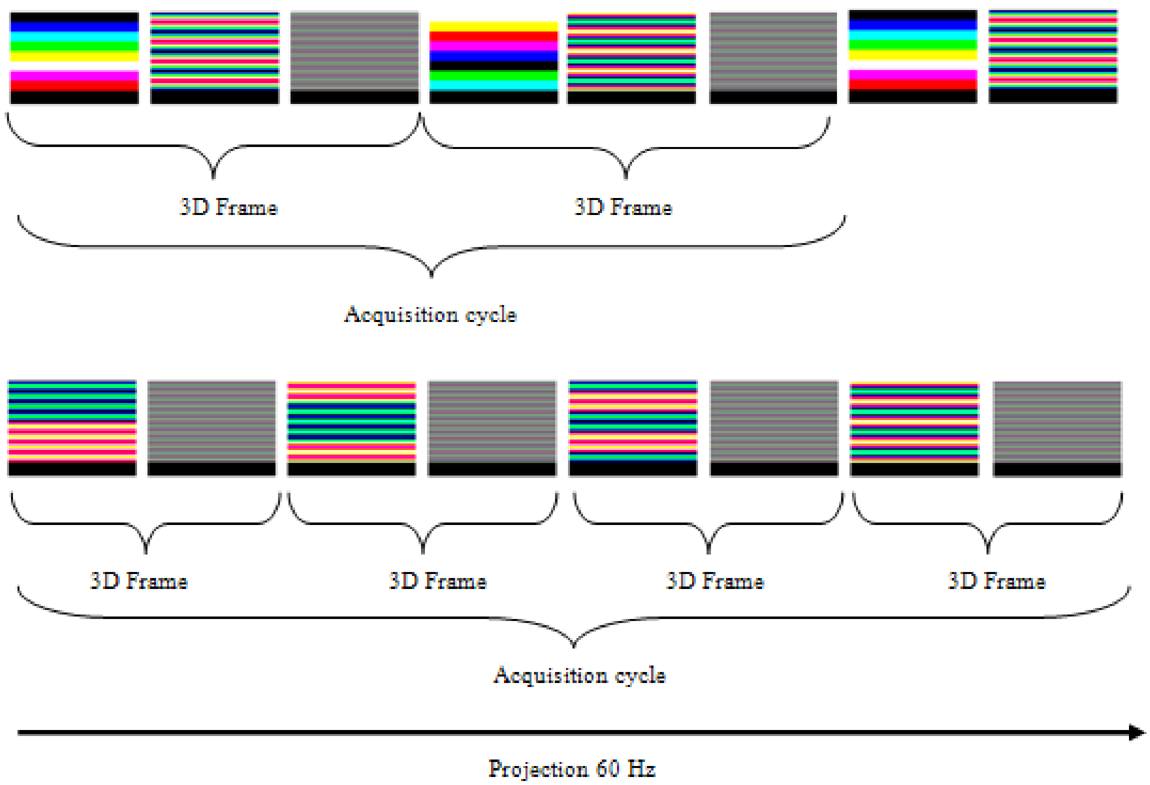

Figure 6. Acquisition cycles: at the top 3 projections are required for a single 3D frame while 2 projections for a single 3D frame at the bottom.

board images the distortion parameters can be accurately estimated, as part of the calibration parameters. We consider only radial distortion and assume tangential ones to be negligible [7, 9]. We use the underlying assumption that straight lines are invariant under projective transformations. Corners can be accurately detected on checkerboard images, and projections of the same $20-50$ lines (with accurately detected corners) of several planes provide a good approximation of a small number of distortion parameters.

After corner detection, the corrected corners are used to find external parameters like rotation, translation, and other internal parameters. The test images of the checkerboard are used to validate the accuracy of the estimated projection matrices.

In our tests, an RMS error of approximately 0.25mm is achieved for objects at a distance of about 70cm from the camera. Projection errors are typically about 0.15 camera pixels and 0.3 projector pixels.

5. Applications

Our high speed color 3D video camera is instrumental in numerous research projects conducted at our research laboratory.

Fig. 7 depicts a sequence of 15 frames (spanning the total of 750 msec) from a scan of a human mouth uttering the sound ”ha” acquired at 20 depth frames per second. The accuracy of the captured facial features and lips and tongue movements make our camera useful for lip reading applications. Other applications of our camera include breast scanning for plastic surgery outcome prediction and visualization (Fig. 8, left); accurate facial scans of non-cooperative newborn infants for diagnosis of chromosomal disorders that have specific facial phenotypic expression (Fig. 8, center); and a 3D facial recognition system capable to accurately acquire faces of moving subjects (Fig. 8, right).

6. Conclusions

We described a cheap and efficient real-time 3D color video capturing device. The proposed 3D camera could be easily and reliably assembled and used by the research and industrial communities. For that purpose, we provide a fully functional model with the supporting software.

In the design of the camera management software, significant emphasis has been made on user-friendly interfaces

Figure 7. A sequence of a real time 3D model zooming on lips and tongue movement.

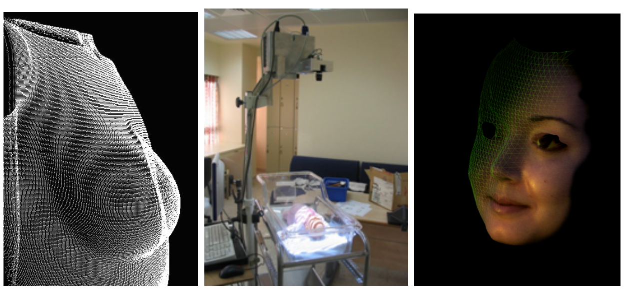

Figure 8. Applications of our 3D camera (left-to-right): aesthetic surgery, new born genetic syndrome identification, and face recognition.

and real-time data acquisition into standard files. Our system supports easy construction of databases for hundreds or thousands of photographed subjects.

The reliability of our system has been demonstrated extensively by numerous laboratory tests during the last six years since the first prototype was built, as well as by several field tests. We have already deployed numerous systems at hospitals and academic institutions. We also found the camera to be robust to moderate changes in lightning conditions.

The main limitation of the system is its limited depth range, though with appropriate optics the latter can be extended from a few centimeters to about ten meters. Low-refresh rate fluorescent light and direct sunlight also reduce the performance of our camera.

References

[1] Dragonfly Technical Reference Manual, 2004. http://www.ptgrey.com/support/. 2

[2] FlyCapture API Programming Reference, October 2006. Version 1.6, http://www.ptgrey.com/support/. 2

[3] PGR IEEE-1394 Digital Camera Register Reference, September 2006. Version 2.0, http://www.ptgrey.com/support/. 2

[4] F. Bernardini and H. E. Rushmeier. The 3d model acquisition pipeline. *Comput. Graph. Forum*, 2(21):149–172, 2002. 1

[5] A. M. Bronstein, M. M. Bronstein, E. Gordon, and R. Kimmel. High-resolution structured light range scanner with automatic calibration. CIS-2003-06 technical report, Technion, 2006. 1, 3

[6] L. Dascal, G. Rosman, and R. Kimmel. Efficient beltrami filtering of color images via vector extrapolation. In *Proceedings of the 1st International conference on Scale Space and Variational Methods in Computer Vision (SSVM 07)*, volume 4485, pages 92–103. Lecture Notes in Computer Science, Springer Verlag, 2007. 3

[7] F. Devernay and O. Faugeras. Straight lines have to be straight: Automatic calibration and removal of distortion from scenes of structured environments. *Machine Vision and Applications*, 13(1):14–24, 2001. 3, 4

[8] I. J. Gavriel, G. Yahav, and O. J. Braun. 3D imaging system, May 2007. http://www.freepatentsonline.com/7224384.html. 1

[9] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, New York, NY, USA, 2003. 3, 4

[10] C. Rocchini, P. Cignoni, C. Montani, P. Pingi, and R. Scopigno. A low cost 3D scanner based on structured light. In *Eurographics 2001*, volume 20, 2001. 2

[11] N. Sochen, R. Kimmel, and R. Malladi. A general framework for low level vision. *IEEE Transactions on Image Processing*, 7(3):310–318, 1988. 3

[12] R. Y. Tsai. A versatile camera calibration technique for high accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal of Robotics and Automation*, RA-3(4):323–344, 1987. 3

[13] M. B. Vieira, L. Velho, A. Sa, and P. Carvalho. A camera-projector system for real-time 3D video. In *DVPR' 05*, 2005. 2

[14] S. Zhang and P. S. Huang. A novel method for structured light system calibration. *Optical Engineering*, 45(8), 2006. 1, 3

[15] Z. Zhang. A flexible new technique for camera calibration. MSR-TR-98-71 technical report, Microsoft Research, 2002. 3