3D single-pixel video

Yiwei Zhang, Matthew P Edgar, Baoqing Sun, Neal Radwell, Graham M Gibson and Miles J Padgett

SUPA, School of Physics and Astronomy, University of Glasgow, Glasgow, G12 8QQ, UK

E-mail: ewayzhang@gmail.com

Received 18 December 2015
Accepted for publication 25 January 2016
Published 29 February 2016

Abstract
Photometric stereo is an established three-dimensional (3D) imaging technique for estimating surface shape and reflectivity using multiple images of a scene taken from the same viewpoint but subject to different illumination directions. Importantly, this technique requires the scene to remain static during image acquisition otherwise pixel-matching errors can introduce significant errors in the reconstructed image. In this work, we demonstrate a modified photometric stereo system with perfect pixel-registration, capable of reconstructing 3D images of scenes exhibiting dynamic behavior in real-time. Performing high-speed structured illumination of a scene and sensing the reflected light with four spatially-separated, single-pixel detectors, our system reconstructs continuous real-time 3D video at ∼8 frames per second for image resolutions of 64 × 64 pixels. Moreover, since this approach does not use a pixelated camera sensor, it can be readily extended to other wavelengths, such as the infrared, where camera technology is expensive.

Online supplementary data available from stacks.iop.org/JOPT/18/035203/mmedia

Keywords: real-time 3D imaging, single-pixel imaging, photometric stereo, Hadamard matrices, pixel registration

(Some figures may appear in colour only in the online journal)
However, reconstruction bias still occurs due to spectral variations for scenes with distinct materials, such as human faces. Here we demonstrate an alternative approach, combining photometric stereo with single-pixel imaging, which utilizes an efficient real-time sampling scheme. Single-pixel imaging [26–28] is a computational imaging technique that allows a single-pixel detector to be used as an imaging device by using a spatial light modulator to provide either time-varying, structured detection of an image or by providing time-varying, structured illumination onto a scene. We have previously shown that when using structured illumination it is the position of the detector determines the apparent lighting condition for the reconstructed image [1]. By using a small number of single-pixel detectors in different spatial locations, multiple images of a scene with different shading profiles can be reconstructed with perfect pixel registration, even for moving objects. By utilizing crossed polarizers it is possible to observe the Lambertian surface reflectivity which allows the surface normals to be estimated via photometric stereo techniques and hence the recovery of 3D images. However, in our original work, many thousands of projected patterns were required leading to acquisition times in the order of several minutes per single image [29, 30].

More generally, within the field of single-pixel imaging, orthogonal and also pseudo-random bases have been employed which significantly reduces the acquisition time, however, the finite modulation rates of micro-electomechanical-systems technology, typically ~20 kHz, places restrictions on the achievable frame rates even for relatively low resolution images. A few studies [31–35] have aimed to improve the imaging speed by using compressive sensing which utilize ‘a priori’ knowledge of the scene, such as sparsity in the spatial frequency domain. Some of the most impressive results utilizing highly-compressed data demand intensive computational processing to recover an image, which does not lend itself well for applications that demand video-rate operation. A variety of alternative compressed sensing schemes have been developed in recent years to enable compressed sensing for large image resolutions. In this work we employ one of these compressive strategies, known as evolutionary compressed sensing [36], in order to demonstrate continuous real-time 3D video at ~8 frames per second for image resolutions of 64 × 64 pixels, equivalent to a speed-up of 4 times compared to a conventional raster-scanning sampling strategy.

**Methods**

**Custom single-pixel system design**

The application programming interface is written as a dynamic-link library file which provides a convenient interference between of control software and accessory light modulator package (ALP) driver. Patterns are first loaded from the controlling software to the ALP board RAM in

---

**Figure 1. Video-rate 3D imaging system.** The system contains a digital micro-mirror device (DMD), a camera lens and four spatially separated photodetectors (PD) fixed surrounding it. A white LED light source is used to illuminate the DMD chip and encoded into binary light fields (0 s and 1 s). A mirror is adjusted manually inside a central 3D-printed mount so that the LED light is reflected to the right direction on the DMD chip. The structured light patterns are projected through the lens onto the object. A plastic polarizer sheet is attached in front of each photodetector (horizontally) and the camera lens (vertically) to eliminate the specular reflection on the object. Both photodetectors and the LED light are controlled by a custom electric board (CEB). Each photodetector receives light scattered by the object to give a signal of intensity value which is then sent to the computer through a data acquisition board (DAB) to form four 2D images. These 2D images are analyzed using a photometric technique to give the 3D information of the scene.
sequence. The display time can also be adjusted manually through the control software, which in this experiment is set to be 50 μs as the minimum. Besides the high speed projection, another important feature of the ALP is that it provides synchronization trigger signals in reference to its display. That is, when one pattern gets displayed, there is a trigger signal released from digital micro-mirror device (DMD) control circuit to the DAQ input which then triggers a series of data acquisition. Signals are acquired after every trigger signal, and the sample numbers are determined by the display time and sampling rate. The DAQ used here is a National Instrument portable USB DAQ (NI USB-6221/16) with a maximum acquisition rate of 250 kHz for all channels. As there are four channels employed, sampling rate for each channel is set to 62.5 kHz. Given that each pattern is displayed for 50 μs, there are approximately three samples acquired for each pattern.

Photometric stereo

The image appearance of an object varies based on the lighting illumination, object orientation, object shape and its reflectance. With a static object, the corresponding surface orientation can be determined by analyzing the object images under different illumination directions. Photometric stereo, which is ideally for Lambertian surfaces, allows depth and surface orientation to be estimated from multiple images of a static object taken from the same viewpoint, but under different illumination directions. The appearance of a diffuse object with a specular varying reflection may be modeled as [37]:

\[
I_p = \sum_{i=1}^{k} \rho'_i f_i(n_p, L_p, \nu),
\]

Where \(I_p\) is the pixel intensity at point \(p\), \(k\) is a fixed value of a linear combination of \(k\) basis materials, \(\rho'_i\) is a reflection
coefficient that varies on the surface, \( f_i \) is any reflectance map as a function of the viewing direction \( v \), \( n_p \) is the surface normal at that point, and \( L_p \) is the incident illumination field.

In general, this method requires those images taken successively following the change of illumination directions. In our system, we replace the lighting sources with single-pixel detectors and the camera with a patterned lighting, in which case those images are acquired simultaneously.

**Basis-scanning with Hadamard matrices**

To maximize sampling efficiency, an orthogonal series of 2D binary patterns, derived from a Hadamard matrix [38], are preloaded to onboard RAM on the DMD. For a Hadamard matrix of order \( 4^k \) (\( k > 0 \), integer), each row is reshaped into a 2D array with \( 2^k \times 2^k \) resolution and upscaled to fill the full height of the DMD, resulting in complete series of \( 4^k \) binary patterns. For example, for complete sampling of a scene at a resolution of \( 32 \times 32 \) would require a series of 1024 reshaped Hadamard patterns to be displayed and the corresponding intensities measured by a photodetector.

**Results**

Before assessing the 3D video quality with our 3D imaging system, we first considered fast 3D single-pixel imaging with a static object. In the system, as illustrated in figure 1, we combined a 3W white LED with a DMD of \( 1024 \times 768 \) pixels, and a camera lens with a 24 mm focal length to deliver structured illumination at a rate of 22 kHz. A data acquisition board (DAB) was used to covert analog intensity measured by four photodetectors into digital signals at a rate of 250 kHz which subsequently were processed to reconstruct both 2D images via computational imaging and a 3D image based on photometric stereo. As used in other work with single-pixel cameras, the Hadamard basis was chosen for providing structured illumination, which yields better quality results compared to raster scanning techniques that suffer from poorer signal-to-noise [37]. In one investigation, we reconstructed a static object (a skull model) at three different resolutions: \( 32 \times 32 \) pixels, \( 64 \times 64 \) pixels and \( 128 \times 128 \) pixels respectively, as shown in figure 2. As anticipated, increasing the image resolution provides improvement to image quality at the expense of the reconstruction frame-rate.
To improve the frame rate we applied compressive sensing algorithms. In choosing the optimal approach we noted that typical images can be represented by a subset of Hadamard patterns instead of a complete pattern set, and continuing adjacent frames are nearly coincident to each other with only slight variations. Hence we ordered the Hadamard patterns based on their corresponding mean signal intensities from the four photodetectors, and utilized the top-ranking patterns to form 2D images. With this compressive sensing algorithm, for each new frame we replaced a small percentage (we chose 10% patterns in this experiment) of low-ranking patterns among the top-ranking patterns with the ones that were randomly selected from the remainder of Hadamard patterns. For each frame, these four 2D images were combined using photometric stereo techniques to obtain 3D images. This evolutionary approach to the selection of a subset of Hadamard patterns is a compromise approach to maintain a high frame-rate without decreasing the spatial resolution.

With implementation of this photometric algorithm, we reconstructed the object at $128 \times 128$ pixel resolution by using five different numbers of pattern pairs: 16,384 pattern pairs, 12,288 pattern pairs, 8,192 pattern pairs, 4,096 pattern pairs, and 2,048 pattern pairs, equivalent to 100% (zero-compression), 75%, 50% 25% and 12.5% compression ratio (see figure 3), and compared the relative root-mean square (rms) errors of the height value in those 3D images to the zero compression result (see table 1). The 3D reconstruction with zero-compression in this figure is the same as the 3D reconstruction of $128 \times 128$ pixel resolution in figure 2. The result shows that, as expected, the rms error of the object’s height value increases when using less pattern pairs (higher compression).

In figure 4, we present a sample of real-time compressed $64 \times 64$ pixel resolution 3D video frames in 1 s. Each 3D frame is produced based on 1024 patterns (25% compression ratio), which equals to the same amount of patterns for a zero-compression $32 \times 32$ pixel resolution 3D image
reconstruction. The frame rate of this 3D video is 7.6 Hz, approximately 4 times faster than the zero-compression 64 × 64 pixel resolution one. In figure 5, we demonstrate a 10 s real-time 3D video frames reconstructed at 128 × 128 pixel resolution at a frame rate of 0.9 Hz, using 4096 Hadamard patterns (25% compression ratio). The object is physically rotated about the z-axis, whilst the 3D reconstructed model is rotated about the y-axis at an angel range of [−30, 30] controlled by the system since we are fully aware of the height map of the object. We notice that the frame rate in this video is restrained due to the fact that the 3D reconstruction process at 128 × 128 pixel resolution starts to play an important role in overall time performance when the 2D reconstruction time decreases with compressive sensing algorithm.

Discussion and conclusion

We have experimentally demonstrated a video-rate 3D imaging system based on photometric stereo and exhibiting perfect pixel registration by utilizing a high-speed structured-illumination and single-pixel detectors. As an extension of our previous related work, we have now shown continuous operation for 3D image reconstruction at image resolutions of 32 × 32, 64 × 64 and 128 × 128 pixels, with Nyquist-sampling frame rates of 8.7 Hz, 2.4 Hz, and 0.5 Hz, respectively. Additionally we have made use of sub-Nyquist sampling (compressive sensing) to speed-up the frame rates for 64 × 64 and 128 × 128 pixel resolution images at the expense of only a modest reduction in image quality, as evidenced by a quantitative analysis. For 64 × 64 and 128 × 128 pixel 3D images, 25% compressive sensing provides increased frame rates of ~8 Hz and ~1 Hz respectively, compared to Nyquist sampling. We note that at 128 × 128 pixel resolution, the total computational time for 2D and 3D image reconstruction, as performed on an octa-core processor, placed a limit on the achievable frame rate, however we anticipate that this may be improved with other processing hardware, such as dedicated high-performance GPU’s. Since this 3D imaging approach does not rely on a pixelated camera sensor and the operational bandwidth of DMD extends beyond this visible, this technique can be applied at wavelengths where 3D imaging is prohibitively expensive.

Acknowledgments

This work was supported by the European Research Council (ERC) through TWISTS (Grant 192382), the Engineering and Physical Sciences Research Council (EPSRC) through COAM (Grant EP/I012451/1) and the University of Glasgow through the Kelvin Smith Scholarship. The data used to produce the content of this manuscript is available at: http://dx.doi.org/10.5525/gla.researchdata.263.

References


Shapiro J H 2008 Computational ghost imaging \textit{Phys. Rev.} A \textbf{78} 061802


