SLI in AR applications

Daniel L. Lau[a] and Barry Ying Yu[b]

[a]University of Kentucky, Address, Lexington, US;
[b]University of Kentucky, Address, Lexington, US

ABSTRACT

This document shows the desired format and appearance of a manuscript prepared for the Proceedings of the SPIE. It contains general formatting instructions and hints about how to use LaTeX. The LaTeX source file that produced this document, `article.tex` (Version 3.3), provides a template, used in conjunction with `spie.cls` (Version 3.3).

Keywords: Manuscript format, template, SPIE Proceedings, LaTeX

1. INTRODUCTION

Structured light illumination is a 3D scanning technique based on projecting a series of striped line patterns and then reconstructing a target surface based on the warping of the stripes as measured by a digital camera.[1] Compared to alternative schemes like passive stereovision[1] or time-of-flight,[1] structured light is far superior in terms of accuracy,[1] but as a triangulation scheme, SLI cannot easily scan at long ranges.[1] Furthermore, as a scanning process, SLI is not typically associated with real-time applications with moving targets.[1]

Scanning time is a major issue in structured light research as numerous papers have been published that focus either on using high speed projector/camera pairs[1] or using unique pattern codecs[1,1,1] that minimize the number of projected patterns. In fact, single pattern SLI techniques do exist[1,1,1] that are quite popular in human computer interfacing applications[1] but which perform poorly in bright ambient light conditions.[1]

With regards to performing structured light scanning with commodity components, a major issue is the inability to properly synchronize a camera and projector such that one can associate a specific pattern to a specific captured image without including some type of key-frame scheme and analyzing the pixels of the captured image. As clearly documented by graphics card manufacturers,[1] GPUs are not real-time systems, and there is no way for the CPU to know when a requested pattern shows up on the HDMI port. And while its possible to guarantee that no particular pattern is dropped, its not possible to guarantee that a particular pattern won't be repeated multiple times in a row.

As a means of addressing one's ability to properly synchronize the camera and projector, there are commercially available projectors that include electronics to embed patterns into the projector which include appropriate GPIO to interface with a camera's trigger;[1,1,1] however, these projectors are inordinately expensive on a per lumens basis. As of the time of this writing, the TI Lightcrafter 4500 retails for $1,250 with only 150 lumens brightness[1] while an LED-based Optoma ML700 has approximately 500 lumens brightness[1] with the same spatial light modulator for $550.[1] Admittedly, the TI Lightcrafter 4500 can project patterns at up to 4,225 Hz, but the 120 Hz of 3D ready projectors is more than adequate for most machine vision cameras limited in frame rate by either the bus to the CPU (USB 3.0, for instance) or by the exposure time required by the sensor.

As a means of ensuring perfect projector/camera synchronization using commodity components, we propose a novel approach to machine vision camera architectures by incorporating HDMI pass-through where the camera reads the incoming HDMI video to extract hand-shaking messages embedded into the top-left pixel. Such handshaking could be used to then trigger the camera to capture the projected frames, ignoring duplicate or don't care frames. Without an incoming HDMI signal, the machine vision camera can alternately produce its

---

Further author information: (Send correspondence to A.A.A.)
A.A.A.: E-mail: aaa@tbk2.edu, Telephone: 1 505 123 1234
B.B.A.: E-mail: bba@cmp.com, Telephone: +33 (0)1 98 76 54 32

own video signal, converting any HDMI projector into an instant structured light scanner which it instantly syncs.

Aside from simple synchronization to a projector, the incorporation of HDMI into a machine vision sensor produces many application possibilities. For instance, cameras could take advantage of EDID[1] to tell a host CPU what kind of camera it is and what its capabilities are including its resolution, color space, and frame rates. Multiple cameras can also by daisy chained together and instantly synchronized together using commodity cables purchased at any big box store with cable lengths of 15, 25, or even 50 feet[1] or longer with appropriate repeaters.[1] And now with the inclusion of ethernet over HDMI,[1] one could easily see GigE cameras converted to HDMI in the near future.

Now an application for our HDMI pass-through cameras that we are particularly interested is projector-based augmented reality such as the AR Sandbox[1] and the AR Pool Table.[1] In the case of the AR Sandbox, a digital projector paired with a PrimeSense RGB+D camera with the RGB+D camera used to monitor the shape of sand in a sandbox placed under the projector as depicted in Fig. **??**. The shape of the sand is then used to generate a psuedo-color relief map which is then projected back onto the sand in perfect register. The result system can then be used to visualize water flow through simulated rivers and streams.[1]

In the AR Pool Table, a digital projector is used to direct the player on where and how to strike the que ball in order to sink the other balls in the various pockets. A machine vision camera monitors the balls' positions on the table as well as the pose of the stick and project augmenting lines onto the table in order to show the player how well their stick alignment matches the alignment posed by the CPU. Working in reverse, the AR Pool Table is now being used to train the PC the play pool using a robotic stick.[1]

In either of the above or other projector-based AR systems,[1,1,1] a common trait of all systems is there inclusion of low-accuracy 3D vision with a commodity projector. Our proposed machine vision camera with HDMI pass through could be instantly inserted between the projector and computer to give that system the option of highly accurate (millimeter) 3D scanning when the RGB+D sensors give only centimeter accuracy. Such high resolution would not only enhance scene understanding in the computer (having highly accurate scans) but also enhance registration between the target surface with the projector as the camera can derive sub-pixel alignment with the projector.

ACKNOWLEDGMENTS

REFERENCES

1. O'Toole, M., Mather, J., and Kutulakos, K. N., "3d shape and indirect appearance by structured light transport," *IEEE Trans. Pattern Anal. Mach. Intell.* **38**(7), 1298–1312 (2016).