Signal Processing: *Image Communication* 40 (2016) 65–81

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

journal homepage: www.elsevier.com/locate/image

A full photometric and geometric model for attached webcam/matte screen devices

Yvain Quéau *, Richard Modrzejewski, Pierre Gurdjos, Jean-Denis Durou

Université de Toulouse, IRIT, UMR CNRS 5505, Toulouse, France

ARTICLE INFO

Article history:
Received 28 February 2015
Received in revised form
11 August 2015
Accepted 14 November 2015
Available online 21 November 2015

Keywords:
Lighting
Camera calibration
Photometric stereo
Extended sources

ABSTRACT

We present a thorough photometric and geometric study of the multimedia devices composed of both a matte screen and an attached camera, where it is shown that the light emitted by an image displayed on the monitor can be expressed in closed-form at any point facing the screen, and that the geometric calibration of the camera attached to the screen can be simplified by introducing simple geometric constraints. These theoretical contributions are experimentally validated in a photometric stereo application with extended sources, where a colored scene is reconstructed while watching a collection of graylevel images displayed on the screen, providing a cheap and entertaining way to acquire realistic 3D-representations for, e.g., augmented reality.

1. Introduction

A lot of common multimedia devices (smartphones, tablets, etc.) are composed of both a screen and a webcam. If this is not the case, some cameras are designed to be easily clipped onto laptops or even monitors. Using such devices, a number of active vision applications attempt to use the camera as a photometric measuring device where the screen is used as a light source. One of the most appealing examples is 3D-reconstruction through photometric stereo [1]: different lightings can be obtained by successively displaying various patterns on the screen [2–7].

Assume that a user is watching a slideshow of images (cf. Fig. 1). A slideshow may correspond to a simple collection of white rectangles with varying locations, as suggested in [2,3,5,6], but also to circular patterns [7], or even to patterns with non-trivial geometry [4], as natural images. Can the light field emitted by the screen be treated as a light source for some applications e.g., 3D-reconstruction through photometric stereo? This question raises the key issue of this paper, which is that of efficiently estimating the light under realistic hypotheses. This introduces two key problems: modelling the emitted light field, and geometrically calibrating the device i.e., determining the camera pose w.r.t. the screen.

The most simple approximation of the screen as a light source is the infinitely distant point light source model. Such a model was considered in uncalibrated photometric stereo algorithms [3,5,7], though the discussion on the resolution of the underlying linear ambiguity (a generalized bas-relief ambiguity if integrability is imposed [8,9]) is rather limited. To avoid such ambiguities, the mean direction and the mean intensity of the light can be calibrated, as do Won et al. in [6]. Still, directional lighting seems rather unrealistic when modelling nearby screens: an anisotropic point light source model is considered instead in [2]. As we shall see in Section 2, this model is physically justified for modelling a single pixel, but does not account for the extended behavior of the screen. It is thus necessary to consider connected sets of pixels, as did Clark in [4] but without considering anisotropy. Our first contribution is to provide a general closed-form expression of the intensity and direction of light emitted by matte screens. As a particular case, we show how to compute the light emitted by a rectangular set of pixels, whatever its

* Corresponding author.
E-mail address: yvain.queau@enseeiht.fr (Y. Quéau).

http://dx.doi.org/10.1016/j.image.2015.11.006
0923-5965/

Fig. 1. Overview of our contributions. While he watches a slideshow of images, a user receives some incident light. In order to use such data in a photometric stereo application (Section 4), we need both to model the light field emitted by the screen (Section 2) and to estimate the pose of the camera (Section 3).

size and location. The light emitted by an image can then be approximated by a straightforward generalization of this model to quadtree-like [10] image decomposition.

The camera/screen calibration requires observing reference points on the screen plane. Due to our device configuration, calibration must be carried out without a direct view of the screen i.e., by using reflections in a mirror. Such a problem has been widely considered in relevant literature [11–16] and recently a solution from a single reflection in one spherical mirror has been reported [17]. In this paper, we aim at describing the most flexible *modus operandi* for users, and at providing a simple algorithm requiring few input data, such that we will consider one planar mirror and a minimal set of three reference points on the screen, as in [15]. Our contribution, initially presented in [18], is to take advantage of the constrained model of the device, as the camera pose only has four degrees of freedom: three for its location and one for its orientation restricted to a rotation around the horizontal axis of the screen. This hypothesis allows us to develop a more efficient calibration method, where as few as two mirror poses and three matched pairs of points are required.

The rest of this paper is organized as follows. After studying in Section 2 the light field emitted by a matte screen, we tackle in Section 3 the camera/screen calibration problem. As an application, we consider in Section 4 the photometric stereo problem using images displayed on the screen as light sources, and show that, using the proposed full photometric and geometric model for the device, reasonably accurate shape and reflectance can be recovered.

2. Light model for matte screens

We start by investigating the light field emitted by the screen: we show that a formal analysis of the problem allows one to reach a highly realistic closed-form model of the emitted light, which only relies on the definition of a Lambertian primary source. After briefly discussing the notion of matte screens, we will introduce a model for the light emitted by a single pixel, considered as an anisotropic point-wise source. Then, this elementary model will be extended into a new theoretical model for planar sources, holding anisotropy, spatially-varying luminance and partial occlusion. Afterwards, we will introduce a framework for simplifying this model when the luminance is uniform and occlusions are ignored. Eventually, we will provide closed-form approximations of the model for rectangular patterns and natural grayscale images, and experimentally validate them on real-world data.

2.1. Matte screens

Let us first introduce the class of monitors targeted in this work: for the sake of simplicity, we only deal with the so-called matte screens (both LCD and LED). Such screens are specifically designed by using an anti-glare coating to limit the apparition of shiny lighting effects, as opposed to bright screens which provide more vivid colors but also stronger reflections. We will also consider “small” viewing angles: when watching the screen from wide angles, brightness obviously tends to be vary much more, even for matte screens. Since we are overall interested in modelling the light emitted towards a user facing the screen, viewing angles can reasonably be assumed to be limited.

The problem tackled here is that of modelling in closed-form the luminous flux emitted by the screen. Considering the screen as a matrix of pixels, the total flux is the sum of the fluxes emitted by all pixels. An experimental study was conducted in [2], where it is demonstrated that the intensity of light emitted by a single pixel (or a small pattern) radially decreases according to a cosine law. We will show in the following that this is actually a consequence of Lambert's law, which states that in the ideal case, brightness must remain constant whatever the viewing point, leading to this anisotropic behavior.

2.2. Case of a single pixel

Our first contribution consists in showing that the empirical model of anisotropic punctual source proposed in [2] to characterize a pixel directly follows from Lambert's law. We define a pixel as a surface element $\mathrm{d}\Sigma_s$ around a point $\mathbf{x}_s = [x_s, y_s, z_s]^\top$ with unit normal $\mathbf{n}(\mathbf{x}_s)$. Let $\mathrm{d}^2\Phi$ denote the amount of luminous flux emitted by the pixel inside the elementary cone of vertex $\mathbf{x}_s$ with solid angle $\mathrm{d}\omega$ and direction $\mathbf{u}_e$ (see Fig. 2). By definition, the *luminance* of the pixel at $\mathbf{x}_s$ is the luminous flux per unit of apparent surface, seen from direction $\mathbf{u}_e$, and per unit solid angle; it is defined by

$$L_{\mathbf{x}_s}(\mathbf{u}_e) = \frac{\mathrm{d}^2\Phi}{\mathrm{d}\omega\,\mathrm{d}\Sigma'_s} \tag{1}$$

where $\mathrm{d}\Sigma'_s = \mathrm{d}\Sigma_s(\mathbf{n}(\mathbf{x}_s)\cdot\mathbf{u}_e)$ denotes the apparent surface of the pixel.

By assuming the pixels to be elementary *Lambertian primary sources*, it follows that $L_{\mathbf{x}_s}(\mathbf{u}_e)$ is independent of $\mathbf{u}_e$, so the luminance will now be referred to as:

$$L_{\mathbf{x}_s}(\mathbf{u}_e) = L(\mathbf{x}_s) \tag{2}$$

Let us now consider an elementary scene surface $\mathrm{d}\Sigma$ around a point $\mathbf{x} = [x, y, z]^\top$ at which the normal is $\mathbf{n}(\mathbf{x})$.