Computer Vision and Image Understanding 145 (2016) 128–138

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

Height from photometric ratio with model-based light source selection☆

William Smith*, Fufu Fang

Department of Computer Science, University of York, York, UK

ARTICLE INFO

Article history:
Received 7 November 2014
Accepted 29 November 2015
Available online 12 December 2015

Keywords:
Photometric stereo
Surface integration
Non-Lambertian reflectance
Albedo estimation

ABSTRACT

In this paper, we present a photometric stereo algorithm for estimating surface height. We follow recent work that uses photometric ratios to obtain a linear formulation relating surface gradients and image intensity. Using smoothed finite difference approximations for the surface gradient, we are able to express surface height recovery as a linear least squares problem that is large but sparse. In order to make the method practically useful, we combine it with a model-based approach that excludes observations which deviate from the assumptions made by the image formation model. Despite its simplicity, we show that our algorithm provides surface height estimates of a high quality even for objects with highly non-Lambertian appearance. We evaluate the method on both synthetic images with ground truth and challenging real images that contain strong specular reflections and cast shadows.

1. Introduction

Photometric stereo has a long history in computer vision [1]. In recent years it has begun to find practical application in areas such as face recognition [2,3], object capture [4,5], medical imaging [6] and surface texture classification [7]. Photometric stereo uses the intensity of reflected light under varying illumination direction to infer the orientation and reflectance properties of a surface. Usually, the only reflectance property that is estimated is the diffuse albedo, although more exotic experimental setups allow estimation of additional reflectance properties such as specular albedo [8], surface roughness [9] and index of refraction [10].

The advantages of photometric stereo are well known: observations are dense (measurements are made at every pixel and resolution is limited only by the resolution of the camera); it can be applied to smooth surfaces devoid of matchable features; estimated surfaces can be relit since both shape and reflectance properties are estimated; and it is able to recover fine scale surface detail. However, surface orientation is only a 2.5D shape representation and the estimated normal field must be integrated in order to recover surface depth or used to refine a 3D mesh captured using other cues [11]. Often, modelling errors mean that the estimated surface normals are subject to low frequency bias leading to distortion in the 3D surface and low global accuracy. Also, photometric methods are usually much more demanding in data capture terms and also in their requirement for controlled conditions.

Despite sustained research effort on the topic, many methods are surprisingly still heavily based on the original approach proposed by Woodham [1]. One of the reasons for this is that Woodham's least squares solution "averages out" errors due to inaccurate assumptions or model inaccuracies. This leads to robust performance even in the presence of significant deviations from the assumptions of Lambertian reflectance with no shadows or specularities, or when there are errors in the light source directions and intensities.

A popular class of approaches that has arisen in the past decade are based on selecting a subset of observed intensities for each pixel [12–15]. We refer to these as *selection-based approaches*. The idea is to exclude observations that are believed to deviate from the modelling assumptions made by the algorithm, for example by excluding shadowed pixels. This transforms the photometric stereo problem into one of labelling pixels according to photometric phenomena. These approaches are motivated by the fact that greater than three observations leads to redundancy which can be exploited by exclusion of noisy data. Almost all of these selection-based approaches rely on ad hoc heuristics that require careful tuning of parameters.

1.1. Related work

The photometric stereo literature is large and here we review only the most relevant work, including selection-based techniques, alternative lighting models and methods that make use of photometric ratios.

The classical approach of Woodham [1] uses each observation to construct a linear equation based on the Lambertian equation (neglecting self shadows). The resulting system of equations is solved

☆ This paper has been recommended for acceptance by Dimitris Samaras.

* Corresponding author.
E-mail addresses: william.smith@york.ac.uk (W. Smith), fangfufu2003@gmail.com (F. Fang).

http://dx.doi.org/10.1016/j.cviu.2015.11.019

in a least squares sense. Hence, a minimum of three observations at each pixel are required to recover the surface normal scaled by the albedo. Woodham's algorithm does not take shadows and highlights into account, they are treated as noise. In practice, this is a strong assumption and it is difficult to image many objects without encountering shadows or highlights. Hence, three image photometric stereo is very sensitive to noise.

Using more than three images results in an overdetermined system and greater robustness to noise. However, systematic noise caused by specularities, shadows, light source attenuation or even a non-linear camera lead to systematic bias in the surface estimates. Since the system of equations at each pixel is independent, Woodham's algorithm also imposes no constraints on surface smoothness or integrability.

The classical approach to this problem is to incorporate smoothness or integrability priors into an energy term expressed as a function of the field of surface normals [16–20]. The drawback of such approaches is that these constraints are only softly satisfied and the result is dependent upon the weight assigned to each prior. There have been some attempts to incorporate photometric constraints into the surface integration process. For example, Chandraker et al. [21] use cast shadows to impose inequality constraints on the surface height estimates.

Selection-based photometric stereo. Barsky and Petrou [12] presented the first selection-based photometric stereo algorithm. Their idea was to use the minimal set of images for which there is redundancy (i.e. four images). A threshold on the reconstruction error using the four image photometric stereo result is used to label pixels where it is believed one observation in the input quadruple is corrupted by the presence of a shadow or specularity. The presence of a highlight is detected using one of two methods. One method is to check if the chromatic distance between the estimated diffuse albedo and the colour of the light sources is above a threshold. Another method is to check if the recovered surface normal could cause specular reflection based on the viewing direction of the camera and the light source direction, i.e. it is close to the perfect specular configuration. If the quadruple does not contain a highlight, it must contain a shadow.

Sun et al. [13] suggested a hierarchical selection strategy for eliminating highlights and shadows. Their method requires images under six illumination conditions and can recover diffuse albedo and surface normals for non-Lambertian surfaces. They first order the observations for a pixel in descending order of brightness. They then check if the first observation contains a highlight, and if the fifth and the sixth observations contain shadows. Highlight and shadow detection is done in the same way as Barsky and Petrou [12].

Hernández et al. [15] proposed a three image photometric stereo algorithm that is robust to shadowing. Their method makes the assumption that for a surface patch, only one pixel from the input triplet is affected by shadow. They use a segmentation algorithm to detect shadowed regions. For the segmented shadowed regions, albedo and surface normal estimates are obtained using only the remaining two observations with integrability being used to resolve the resulting ambiguity.

A number of approaches can be viewed as outlier removal. Miyazaki et al. [22] compute surface normals from all possible observation triplets and then compute the median direction as a robust estimate of the surface normal. Yu et al. [23] propose an algorithm based on the maximum feasible subset framework. The idea is to select the maximum subset of observations that satisfy the Lambertian constraint. Mukaigawa et al. [24] used a random sampling approach to eliminate non-Lambertian observations. These approaches are related to our selection strategy, in the sense that we also detect and remove outliers. The difference is that our notion of outliers is based on deviation from predicted appearance using an initial (possibly naive) model whereas theirs are based on statistical analysis of the input data.

Ikehata et al. [14] posed the problem of selection as imposing sparsity on a Lambertian error matrix. However, strictly imposing sparsity leads to a non-convex optimisation problem. While not selection per se, Higo et al. [25] use the idea of "consensus" in photometric stereo. Namely, observations merely reduce the size of the solution space for a given surface normal. This allows them to use very general models, making relatively weak assumptions about reflectance and camera properties. However, their approach is data-heavy, requiring very large numbers of images to build a sufficient consensus on the correct normal direction. Moreover, cast shadows cannot be modelled and act as noise in the consensus process.

Rather than make per pixel observation selections, recently, Argyriou et al. [26] tackled the problem of choosing the *globally* optimal configuration of light source directions. The idea is that for a class of objects, e.g. faces, there is likely to be a configuration of light sources that minimises the number of shadowed observations over the whole image set on average. They pose the problem of searching for the lighting configuration as one of sparse optimisation.

Alternative illumination models. When a surface is lit by a point source (and assuming no inter-reflections), shadowing is a binary function, i.e. a point on the surface is either lit or it is shadowed. Under a continuous field of illumination, this is not the case as some portion of the illumination may be visible and observations need not be selected and excluded in a binary fashion. The analogue of shadowing under continuous illumination is occlusion, where part of the local upper hemisphere is occluded by other parts of the surface.

The special case of a uniform (i.e. ambient) continuous field of illumination was studied by Prados et al. [27] who demonstrated that shading under such illumination can be ambiguous. Allowing arbitrary illumination but neglecting occlusions, the most general approach is due to Basri et al. [28]. They use a spherical harmonic model of Lambertian appearance to develop a photometric stereo algorithm for complex illumination environments. More recently, such a setting has been considered in a single image, shape-from-shading setting. For example, Huang and Smith [29] provide a linear approach for shape recovery under a first order spherical harmonic lighting model.

Rather than considering natural illumination environments, Ma et al. [8] developed a variant of photometric stereo which operates under continuous linear gradient illumination fields. They pose surface normal recovery as estimation of the centre of mass of the reflectance lobe. To achieve this, they compute the first moment of the reflectance function by integration with a linear gradient in each direction of the coordinate system. Although they do not model occlusions, their approach degrades gracefully with increasing occlusion. In fact, their method estimates the "bent normal" direction often used in graphics (the mean unoccluded direction).

Another relaxation of assumptions is to allow near-field (i.e. local) lights. Here, the attenuation with distance must be modelled. Recently, Papadhimitri and Favaro [30] showed that uncalibrated near-field photometric stereo suffers from reduced ambiguities in comparison to the distant case.

Photometric ratios. The idea of using ratios between photometric measurements to aid surface analysis was first proposed by Davis and Soderblom [31]. In contrast to most later work, they used intensity measurements from a single image. By finding profiles with assumed equal topographic and albedo variations, they take ratios to cancel out surface reflectivities.

The use of ratios between images under two different illumination conditions was first proposed by McEwen [32] in the context of remote sensing. Lee and Brady [33] were the first to exploit the resulting constraint in a computer vision context. Specifically, they took ratios between observations from different viewpoints leading to an equation in one of the components of the surface gradient. They use this to estimate depth which they subsequently combine with depth estimates from binocular stereo. This was followed by another hybrid photometric/geometric shape estimation method from Wolff and