AUTHOR QUERY FORM

ELSEVIER	**Journal:** COMIND **Article Number:** 2453	**Please e-mail or fax your responses and any corrections to:** **E-mail:** corrections.esch@elsevier.thomsondigital.com **Fax:** +353 6170 9272

Dear Author,

Please check your proof carefully and mark all corrections at the appropriate place in the proof (e.g., by using on-screen annotation in the PDF file) or compile them in a separate list. Note: if you opt to annotate the file with software other than Adobe Reader then please also highlight the appropriate place in the PDF file. To ensure fast publication of your paper please return your corrections within 48 hours.

For correction or revision of any artwork, please consult http://www.elsevier.com/artworkinstructions.

Any queries or remarks that have arisen during the processing of your manuscript are listed below and highlighted by flags in the proof. Click on the 'Q' link to go to the location in the proof.

Location in article	Query / Remark: click on the Q link to go Please insert your reply or correction at the corresponding line in the proof
Q1	Please confirm that given names and surnames have been identified correctly.
Q2	Please check the telephone number of the corresponding author, and correct if necessary.
Q3	Fig. 1 will appear in black and white in print and in color on the web. Based on this, the respective figure caption has been updated. Please check, and correct if necessary.
Q4	Please provide complete details for Ref. [15].
Q5	Please supply the name of the city of publication for Ref. [47].
	Please check this box or indicate your approval if you have no corrections to make to the PDF file ☐

Thank you for your assistance.

Computers in Industry xxx (2013) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Computers in Industry

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

Robust 3D face capture using example-based photometric stereo

Rafael F.V. Saracchini [a,*], Jorge Stolfi [a], Helena C.G. Leitão [b], Gary A. Atkinson [c], Melvyn L. Smith [c]

[a] Institute of Computing, State University of Campinas, Brazil
[b] Institute of Computing, Federal Fluminense University, Brazil
[c] Machine Vision Laboratory, University of the West England, United Kingdom

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Photometric stereo
Shape-from-shading
3D surface reconstruction
Facial geometry capture
Image processing

ABSTRACT

We show that using example-based photometric stereo is possible to achieve realistic reconstructions of the human face. The method can handle non-Lambertian reflectance and attached shadows after a simple calibration step. We use spherical harmonics to model and de-noise the illumination functions from images of a reference object with known shape, and a fast grid technique to invert those functions and recover the surface normal for each point of the target object. The depth coordinate is obtained by weighted multi-scale integration of these normals, using an integration weight mask obtained automatically from the images themselves. We have applied these techniques to improve the PhotoFace system of Hansen et al. (2010) [10].

1. Introduction

Capture of the three-dimensional geometry of the human face is a computer vision problem with many applications, including facial recognition and mood detection, computer animation, plastic surgery and automated sculpture. In particular, it has been shown in recent years that the use of the 3D geometry of the face improves the robustness of facial recognition methods under variations in illumination, pose and perspective [1–4].

Several techniques have been used for 3D facial geometry capture, including laser ranging [5], structured lighting [6], geometric stereo [7] and shape-from-shading [8]. Shape-from-shading (with a single illumination) is limited to objects with uniform color and finish, and is too unreliable for practical facial recognition. Currently the only methods that are being used in commercial systems are laser ranging and geometric stereo, with or without structured lighting. However these methods often require the target to remain still during the scanning, and require bulky and expensive specialized equipment. These factors significantly limit their application.

Morphable models [9] have been proposed as a way to handle variations in illumination, pose and perspective without full 3D geometry capture. However, those methods are inherently limited in their accuracy and robustness under perturbations, like glasses or facial hair that are not previously included in the morphable model; and are unable to capture fine details of skin. They also have a somewhat high computational cost, which clearly limits application.

Variable lighting photometric stereo, here called simply *photometric stereo* (PS), is the extension of shape-from-shading using multiple images with a single viewpoint but different illumination conditions. As a potential 3D facial capture method, its main advantages are that it requires very simple and inexpensive equipment, can be used in ordinary environments without hampering the subject's motion or demanding his cooperation, and can capture high-resolution 3D data in a fraction of a second [10–12]. Indeed, Broadbent et al. demonstrated a photometric stereo system that captures 640×480 depth maps at video rates (15 frames-per-second) using a PC with a popular graphics card [13].

Photometric stereo does not obtain the depth information directly; instead it measures the average normal of the surface within each image pixel. The slope data is then integrated to provide the relative depths of different parts of the object. These integrated depths are somewhat less accurate than those obtained with laser ranging and structured light, but are accurate enough for facial recognition, considering the natural variation of human face geometry over time. Indeed, the slope data can be used directly in facial recognition algorithms [14,15].

Photometric stereo requires information on the finish of the surface and on the illumination conditions of each input image. Specifically, it needs for each image the *shading function* that maps a surface normal to the corresponding shading factor. It also needs a *weight map*, a mask that identifies the image pixels where the

* Corresponding author. Tel.: +55 2191975833.

0166-3615/$ – see front matter © 2013 Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.compind.2013.04.003

brightness data is not reliable or relevant for some reason (e.g. for being part of the background, for being affected by projected shadows, or for straddling silhouette edges of the object). This mask is essential for accurate integration of the slope maps [16].

Here we consider specifically *example-based* photometric stereo (EBPS), a variant of the method where the shading functions are obtained by analyzing images of an *example object* with known geometry and the same finish as the scene, under the same illumination conditions. The main contributions of this paper are (1) the use of high-order spherical harmonics to model the shading functions and remove noise from the example images; (2) an unsupervised method to construct the weight mask of each image; and (3) the use of these methods to improve the PhotoFace real-time face capture system of Hansen et al. [10].

2. Basic concepts and notation

The basic problem of photometric stereo is to determine the orientation of the surface (that is, the surface normal) at every visible point of an opaque object, given $m \geq 3$ digital images $S_1, \ldots, S_m$ of it, all taken with the same pose and viewpoint but with distinct illumination conditions. We assume that the images are geometrically aligned and photometrically corrected, so that they have a common domain $\mathcal{S} \subseteq \mathbb{R}^2$, and that the samples $S_1[p], \ldots, S_m[p]$ at the same point $p \in \mathcal{S}$ are the apparent radiances of same point $P[p]$ on the visible surface of the target object, under the various lightings. The goal is then to determine surface's normal $\vec{s}[p]$ at that point.

For simplicity we assume that the images are monochromatic. We also assume that the camera's field of view is sufficiently narrow and the light sources in each image are sufficiently far away that the lighting and viewing direction can be assumed to be uniform over the entire target object. The method can however be extended to color images, non-uniform light fields, and conical instead of parallel image projection.

In the simplest version of the photometric stereo problem, one also assumes that the surface's finish is known, isotropic and uniform. More precisely, one assumes that the surface's bidirectional radiance distribution function (BRDF) $\sigma[p]$ at $P[p]$ is the product of a known isotropic and absorption-free BRDF β, the *surface finish*, and an unknown factor $\hat{s}[p] \in [0,1]$, the *albedo* at that point (also called the *intrinsic color* or *light absorption coefficient*). It follows that the intensity $S_i[p]$ of each image pixel can be analyzed as the product of the albedo $\hat{s}[p]$ and a *shading factor* that depends only on the image index i and on the surface's normal $\vec{s}[p]$ at that point. Specifically,

$$S_i[p] = \hat{s}[p]\, L_i(\vec{s}[p]) \tag{1}$$

Here, each L_i is the *shading function* for image S_i, that maps a unit vector $\vec{n} \in \mathbb{S}^2$ to the apparent radiance of a white surface with BRDF β, oriented with normal $\vec{n}$ under the lighting conditions of image S_i. The shading functions are related to the finish BRDF β by the equation

$$L_i(\vec{n}) = \int_{\mathbb{S}^2} \Phi_i(\vec{u})\beta(\vec{n},\vec{u},\vec{v})\, d\vec{u} \tag{2}$$

where $\vec{v}$ is the viewing direction (from the point $P[p]$ toward the camera), and $\Phi_i(\vec{u})$ is the intensity of the light flow in the direction $-\vec{u}$ (that is, the radiance of the "sky" in the direction $\vec{u}$) prevailing in image S_i. If the illumination for image i was provided by a distant point source in the direction $\vec{u}_i^*$, the integral reduces to

$$L_i(\vec{n}) = \Phi_i^*\beta(\vec{n},\vec{u}_i^*,\vec{v}) \tag{3}$$

where the factor Φ_i^* quantifies the intensity of that light source. Note that we are including the geometric factor $\max\{\vec{n}\cdot\vec{u},0\}$ in the finish BRDF β, and we use these equations only when $\vec{n}\cdot\vec{v} > 0$. Note also that, by the uniform lighting assumption, the shading function L_i does not depend on the position p, except through the normal $\vec{s}[p]$.

This lighting model allows attached shadows, and is adequate for scenes consisting of a single mostly convex object. On the other hand, this model cannot account for projected shadows, radiosity effects, or sources with uneven light distribution.

With these definitions, we can formally state the basic problem of photometric stereo as follows: given the intensities $S_0[p], \ldots, S_m[p]$ for a pixel p, find the normal vector $\vec{n}[p]$ and the albedo $\hat{s}[p]$ that satisfy Eq. (1) for all i. This problem generally can be solved if the illuminations are sufficiently varied and the finish BRDF β is dominated by wide-angle scattering: that is, more like the Lambertian BRDF than that of a mirrored or glossy black surface.

2.1. Observation vectors

The *observation vector* of a pixel $p \in \mathcal{S}$ is the m-vector of its radiances in all the images, that is,

$$\boldsymbol{S}[p] = (S_1[p], S_2[p], \ldots, S_m[p]) \tag{4}$$

We define the *shading vector function* as the list of all m shading functions, that is, the function $\boldsymbol{L}$ from $\mathbb{S}^2$ to $\mathbb{R}^m$ such that

$$\boldsymbol{L}(\vec{n}) = (L_1(\vec{n}), L_2(\vec{n}), \ldots, L_m(\vec{n})) \tag{5}$$

The basic problem of photometric stereo can be stated more succinctly as: given the vector $\boldsymbol{S}[p]$ of a pixel p, find the unit vector $\vec{n} \in \mathbb{S}^2$ such that the vector $\boldsymbol{L}(\vec{n})$ is practically collinear with $\boldsymbol{S}[p]$, that is, the angle between them is nearly zero. (Needless to say, the inevitable measurement errors in the radiances $S_i[p]$ will introduce random perturbations in the observation vector $\boldsymbol{S}[p]$, so one cannot expect exact collinearity.) Then we can infer that the surface normal $\vec{s}[p]$ at p is $\vec{n}$, and that the albedo $\hat{s}[p]$ is the ratio $\|\boldsymbol{S}[p]\|/\|\boldsymbol{L}(\vec{n})\|$ between the m-dimensional Euclidean norms of the two vectors.

2.2. Observation signatures

We can remove the albedos from the problem by normalizing the observation vectors and the shading vector function. Namely, we define the *observed signature* $\boldsymbol{s}[p]$ of a pixel p as being its observation vector normalized to unit length; and the *shading signature function* $\boldsymbol{l}$ as the shading vector function normalized in the same way. That is,

$$\boldsymbol{s}[p] = \frac{\boldsymbol{S}[p]}{\|\boldsymbol{S}[p]\|}, \quad \boldsymbol{l}(\vec{n}) = \frac{\boldsymbol{L}(\vec{n})}{\|\boldsymbol{L}(\vec{n})\|} \tag{6}$$

Note that $\boldsymbol{s}[p]$ is a vector on the sphere $\mathbb{S}^{m-1}$, and $\boldsymbol{l}$ is a function of $\mathbb{S}^2$ to $\mathbb{S}^{m-1}$. Then the photometric stereo problem reduces to computing the functional inverse of the shading signature function; that is, find $\vec{s}[p] \in \mathbb{S}^2$ such that $\boldsymbol{l}(\vec{s}[p])$ is as close as possible to $\boldsymbol{s}[p]$ in the norm $\|\cdot\|$ (which is a monotonic function of the angle between the vectors).

2.3. The sufficient data hypothesis

The photometric stereo approach will fail if the shading signature function is not invertible, that is, if there are two normal directions $\vec{n}', \vec{n}''$ with collinear shading vectors $\boldsymbol{L}(\vec{n}') = \alpha\boldsymbol{L}(\vec{n}'')$ for some scalar α. To avoid this problem, the illumination conditions must be sufficiently varied to break any such ambiguities. In particular, the light sources used in all images must not be all in the same plane, and every visible point of the target object must be illuminated on at least three of the images. We will assume that these conditions are satisfied in what follows.

On the other hand, example-based PS can in principle work with non-Lambertian BRDFs and arbitrary lighting, as well as with *attached* shadows and penumbras, since these effects do not destroy the proportionality between the vectors $\boldsymbol{S}[p]$ and $\boldsymbol{L}(\vec{n})$. In particular, there is no need to identify the images and pixels where attached shadows occur. (*Cast* shadows and scene-scattered light, however, are still a problem.)

3. Related work

Variable-lighting photometric stereo was first studied as a computer vision problem by Woodham in the late 1980s [17]. In his pioneer work, Woodham demonstrated that is possible to recover the inclination of the surface using at least 3 non-coplanar point-like light sources.

Shadows and glossy reflections: In the following years, Woodham's results were improved and extended in many ways. Woodham originally assumed that the scene had Lambertian finish, and considered only points that were fully illuminated by all three sources with known directions. In that case the surface normal vector can be computed from the three intensity values of the pixel by a simple analytic formula. Barsky and Petrou [18] later showed how to handle glossy highlights and shadows using only 4 images, provided that such anomalies affect the value of a pixel in at most one of the four images. Yu et al. [19] then used linear programming to extend this result for an arbitrary number of lightings, allowing multiple anomalies in each pixel. Due their simplicity and low processing cost, these methods are now commonly used for real-time and fast capture applications [11,13].

Unknown lighting: In practice one often faces a more difficult version of the problem where the lighting conditions of each image are not known a priori and must be determined from the images themselves. Hayakawa [20] addressed a limited version of this problem, assuming that each image was illuminated by a distant point source with unknown direction. He claimed that the normal directions could be obtained through singular-value decomposition of the input data, viewed as matrix where each row is a pixel and each column an input image. This solution was later improved by Yuille et al. [21,22]. Basri et al. [23,24] further generalized the solution using spherical harmonics to handle ambient and semi-diffuse lighting. However, Hayakawa's approach is limited because of inherent ambiguities in the problem, and rather sensitive to the arrangement of light sources. The resulting normals are affected by indeterminate factors that must be determined by other criteria and explicitly corrected for.

Unknown surface finish: Aldrin et al. [25] and Goldman et al. [26] considered the more difficult problem where the surface finish β is unknown. They assumed that β of the surface was some unknown linear combination of a finite library of "model" BRDFs (Lambertian, glazed, etc.), with different coefficients in each pixel. McGunnigle et al. [27] were able to recover good estimates of surface of metallic objects under very strict lighting conditions. Higo [28] does not attempt to model the BRDF, assuming only that it is monotonic and isotropic and achieves good results in presence of glossy highlights and shadows. These methods require dozens of input images, and their high computational cost prevents their use in real-time applications.

Example-based photometric stereo: In spite of the advances described above, the problem of photometric stereo with an unknown finish BRDF and/or unknown lighting remains fraught with practical and computational difficulties. The example-based approach to photometric stereo sidesteps these difficulties by extracting the shading functions L_i directly from m calibration images $G_1, \ldots, G_m$ of a reference object with known shape and albedo, taken from the same viewpoint and under the same lightings as the images $S_1, \ldots, S_m$.

This approach was introduced by Woodham in 1989 [17] and further explored recently by Herztmann and Seitz [29,30] and by us [31] The inconvenience of having to place a reference object in the scene is compensated by the fact that recovery of the surface normals with very complex BRDFs and arbitrary lighting is possible. Moreover, the shading functions L_i, are inherently smoother (and therefore easier to model) than the finish BRDF β and the light flows Φ_i. The example-based approach is also very efficient and usually produces good results with only 6–12 images.

3.1. Slope integration

In the context of this paper, "3D geometry capture" means determining for each pixel $p \in \mathcal{S}$ the height $z[p]$ of the surface visible in p, relative to some arbitrary reference plane perpendicular to the viewing direction $\vec{v}$. Photometric stereo does not yield directly this information, but only the surface normal vector $\vec{s}[p]$ within that pixel. The normal can be trivially converted to the *height gradient* or *slope* vector $\vec{\nabla} z[p]$, the vector of $\mathbb{R}^2$ consisting of the X and Y derivatives of the height map z. The gradient map $\vec{\nabla} z$ then must be *integrated* to yield the height map z.

The integration of gradient maps obtained by photometric stereo is not a trivial task. For one thing, the gradient data is not continuous, but *discretized*, that is, available only in the center of pixels, forming a regular orthogonal grid. Moreover, the gradient data $\vec{\nabla} z[p]$ is contaminated by errors caused by camera noise, violations of the photometric stereo assumptions (opaque surface, uniform lighting, constant finish, etc.) and approximations in the photometric stereo algorithm itself (such as the error due to the use of a finite table to invert the shading signature function).

For some pixels, the magnitude of such perturbations may be so high that the gradient value $\vec{\nabla} z[p]$ is practically unknown. That happens, in particular, in background areas that are outside the range of the light sources, or where the image is very dark, or where the surface is covered by hair or other non-trivial 3D texture, or where the pixel straddles a discontinuity in the height function, such as a silhouette edge (the boundary of the projection of some foreground object). One should note that the gradient data $\vec{\nabla} z[p]$ for a pixel p is a non-linear function of the pixel radiances $S_1[p], \ldots, S_m[p]$; and each $S_i[p]$ is the average of the radiance of the surface over some finite region corresponding to p, which in turn is a non-linear function of the surface's true gradient. Therefore, if the true surface gradient varies considerably within the pixel, the gradient $\vec{\nabla} z[p]$ computed by photometric stereo may be substantially incorrect.

These unavoidable errors in the photometric gradient $\vec{\nabla} z$ require the use of specialized integration algorithms. A comparative survey was presented by us in a previous article [16]. Some integrators that are still widely used in this problem, such as path integration [32] and Frankot–Chellapa's Fourier-based method [33] simply ignore those errors, and may produce very incorrect height maps. See Figs. 1 and 2.

Weighted Poisson integrators: The only integrators that can cope with missing and unreliable gradient data are the weighted Poisson-based methods [34,35,16,36]. Those methods require an extra input, a *weight map* that specifies the reliability $w[p]$ of each gradient value $\vec{\nabla} z[p]$, as a number between 0 ("meaningless") and 1 ("maximally reliable"). Those integrators set up a system of linear equations that relate the given gradients $\vec{\nabla} z[p]$ to finite differences of unknown heights $z[p']$ of adjacent pixels. The linear system is then solved to obtain the heights.

Fast system solving: The linear system built by Poisson integrators is sparse but very large: it has one unknown height and one equation for each pixel with non-zero weight, and each equation typically has five to nine non-zero coefficients. Solving it by Gauss elimination requires super-linear space; solving it by Gauss–Seidel iteration requires way too many iterations. We use a

Fig. 1. A simple height map (left), its height gradient map with some noise added (middle) and the recovered height map (right) produced from the gradient by the Fraile–Hancock tree-path integrator [32]. (In the gradient map, the *X* and *Y* derivatives are combined into color values.) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

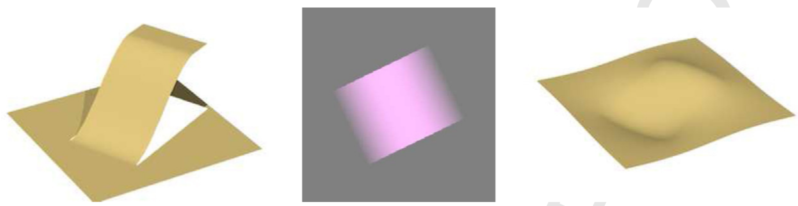


Fig. 2. A height map with discontinuities (left), its height gradient map (middle) as it would be computed by photometric stereo without any noise, and the recovered height map (right) computed from the gradient by the Frankot–Chellapa Fourier-based integrator [33].

multi-scale iterative solver that we developed specifically for this problem that uses a linear amount of memory and runs in linear time for typical instances [16].

3.2. Photometric stereo for 3D face capture

Due to its non-intrusive nature and lower cost, photometric stereo has received considerable attention in recent years as a method for the capture of facial geometry that may be optimal for certain applications. Its viability was demonstrated by Yuille et al. [22], using their proposed SVD method for face images captured under light sources with unknown direction. However, their method demanded a large number of images to overcome errors introduced by deviations of the Lambertian reflection model. Georghiades et al. [37] managed to improve this result for human faces using only 7 distinct light sources, by discarding values that were under or over a predetermined threshold. Lee et al. [38] extended his approach, using 3 known light sources and an arbitrary number of images with unknown lighting.

The PhotoFace system built by Hansen et al. [10] demonstrated the use of photometric stereo for nearly instantaneous face capture of people in motion. This system is described in more detail in Section 4. It uses four point-like light sources, and computes the normals by a simple analytic method that assumes a Lambertian surface finish. The normals are used as input for face recognition algorithms.

A major difficulty in this application is that skin has a complicated non-Lambertian finish, due to its structure of multiple translucent layers.

Real-time face capture: The capture of facial 3D geometry in real time, possibly at video frame rates, has received significant attention in the last few years. For greater speed, some of these systems use multi-spectral method, namely color cameras and colored lights, to capture three or more images S_i at the same time. This technique too was pioneered by Woodham, in 1994 [39].

Using a modified version of Hayakawa's SVD method, Schindler [40] achieves real-time capture using a color monitor as a multi-spectral the light source. The captured geometry is not very accurate, being intended to be used in face modeling and video-chats. Vogiatzis et al. [12] uses a combination of shape-from-motion and multi-spectral photometric stereo, allowing for glossy reflections using a modified Phong model. Fyffe et al. [41] uses a customized 6-channel camera to recover albedo and surface inclination with a single picture.

4. The PhotoFace system

Our work builds on the PhotoFace project, a hardware and software system developed in 2010 by Hansen, Atkinson, and others at the Machine Vision Laboratory (MVL) of the University of the West of England (UWE) [10]. PhotoFace was designed for automatic, near-instantaneous, non-intrusive 3D face capture of people walking through an instrumented booth. Its demonstration prototype consisted of a aluminum frame structure with a high-speed photographic camera and four near-infrared light sources. The system was triggered by an ultrasound sensor as the Person was about 2 m away from the camera and walking toward it. The camera snapped four photos of the person in quick succession, while each light was flashed in turn, and then one more with all lights turned off to record the ambient light. See Fig. 3.

4.1. Specifications

All PhotoFace devices are controlled by a standard PC. The light sources and the ultrasound trigger are connected to the computer

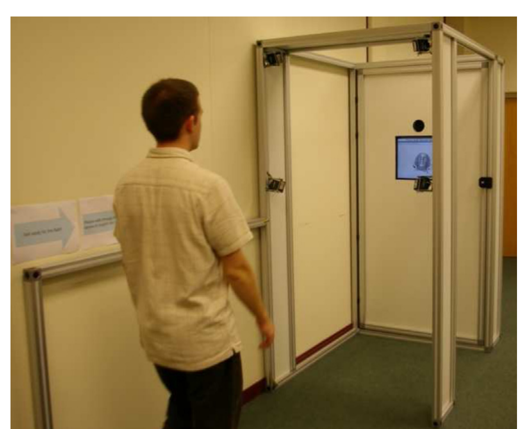


Fig. 3. The PhotoFace prototype built at UWE MVL.

Please cite this article in press as: R.F.V. Saracchini, et al., Robust 3D face capture using example-based photometric stereo, Comput. Industry (2013), http://dx.doi.org/10.1016/j.compind.2013.04.003

via an NI PCI-7811 DIO card. An NI PCIe-1426 frame grabber is also connected to the DIO card via a RTSI bus for triggering purposes, and to the camera via a Camera Link© interface. The latter is used to send the triggering signal to the camera and to store the frame data in the computer. The camera is also directly connected to the DIO card. All interfacing code is written in NI LabVIEW and coordinates all the image capture sequence.

Light sources: Each of the four near-infrared light sources is a VIS080IR 7-LED cluster, which emits light at ≈850 nm. The four lights are arranged in a irregular rectangular shape pointing toward the general area where the person will walk through. This disposition ensures that the light source directions would not be co-planar, in order to obtain good quality images for each illumination triplet.

Camera: The camera is a Basler 504 kc model with a 55 mm, f5.6 Sigma lens positioned in the center of the rectangle formed by the light sources. It is able to capture an 1280 × 1024 8-bit RGB image in less than 5 ms.

Trigger: The system's trigger is an ultrasound sensor, a highly directional Baumer proximity switch that is activated when its beam is broken within a distance of 70 cm.

4.2. Image capture

To ensure the alignment of the images (required by photometric stereo methods), the capture should be performed in a very small time interval. It was verified experimentally that is necessary at least a frame rate of 150 frames-per-second and a accurate synchronization of light sources.

The frame grabbing process is started when a person crosses the beam of the ultrasound sensor. Upon receipt of the sensor's signal, the DIO tells the camera to start the frame integration process. Once the camera confirms that the integration has started, the DIO card flashes one of the lamps, and waits for the camera to signal the end of frame capture. The DIO card repeats this sequence for the other three lights, and then once more with all lights turned off. The entire capture process takes about 50 ms, which seems to be instantly to a human observer.

4.3. Image processing

The "dark" image is subtracted from the four illuminated images to remove the contribution of ambient lighting. Since the subjects may vary substantially in height, the camera is set up to capture an area many times larger than the person's face. The images are therefore cropped to the face's approximate bounding box (typically ≈400 × ≈500 pixels), which is determined by the algorithm of Lienhart and Maydt [42]. See Fig. 4.

As described by Hansen et al., the PhotoFace software used an analytic method to compute the surface normal at each pixel. The method assumed a Lambertian surface finish and a single distant point-like light source in each image. First, their algorithm computed a tentative normal $\vec{s}'[p]$ by assuming that the pixel was illuminated by all four sources. Under this assumption the shading Eqs. (1)–(3) reduce to an overdetermined system of 4 equations on 3 unknowns, that was solved with the least squares criterion.

As a check for the presence of shadows, they then computed a second normal $\vec{s}''[p]$ by the same method but excluding the image S_i with smallest $S_i[p]$. If $\vec{s}''[p]$ made an obtuse angle with the direction of the excluded light source, they assumed that the point was in that light's shadow and set $\vec{s}[p]$ to $\vec{s}''[p]$, otherwise they set $\vec{s}[p]$ to $\vec{s}'[p]$. Height maps were then computed from the normals using the Frankot–Chellapa integrator [33]. See Fig. 5.

Calibration: The analytic method used required the direction of each light source relative to the person's face. Since the light sources were not permanently fixed to the frame, a calibration step was performed once after assembly to obtain that information. For that purpose, a capture sequence was performed with a reflective sphere in place of the person's face. The direction of each light source was then determined by the location of its reflection (a

Fig. 4. Four images captured by PhotoFace after subtraction of ambient lighting and trimming by the face detection algorithm.

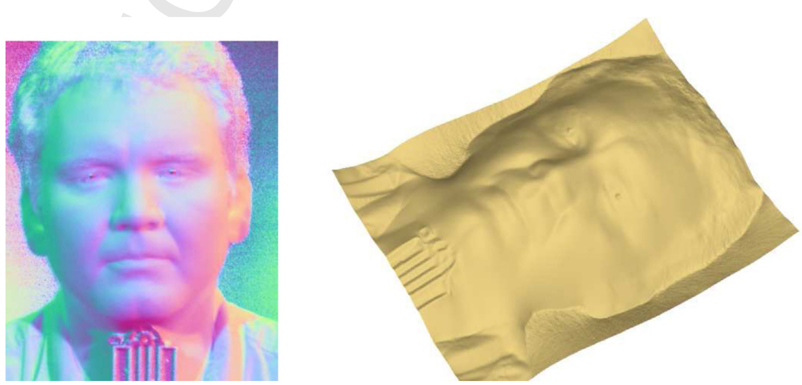


Fig. 5. Normal and height map computed from the images of Fig. 4 by the PhotoFace processing pipeline.

Please cite this article in press as: R.F.V. Saracchini, et al., Robust 3D face capture using example-based photometric stereo, Comput. Industry (2013), http://dx.doi.org/10.1016/j.compind.2013.04.003

bright spot) on the sphere. The four light sources were assumed to have the same intensity.

4.4. Our improvements

The analytic method used in the first version of PhotoFace was very fast, but since it assumed a Lambertian finish for the skin it yielded rather inaccurate normals wherever the image was affected by glossy reflections. The method also failed usually in those regions (mainly under the chin and nose) which were illuminated by only two of the four lights.

The Frankot–Chellapa integrator too was chosen for its speed: it took only a few seconds to process each normal map. However, since it uses the Fast Fourier Transform algorithm, it necessarily gives equal weight to all pixels. Therefore, the integrated face geometry was often distorted by the spurious gradient values in the background and shadowed areas. These distortions prevented the use of the height maps for face recognition; the slope maps were used instead [10].

In the remainder of this article we describe an improved image processing system we developed for the PhotoFace hardware. Namely, we replaced the analytic normal computation algorithm by the example-based method, which can in principle handle the semi-glossy finish of human skin. For this purpose we developed a reliable method to extract the shading function from the images of the example object that removes most of the noise present in those images. This method is described in Section 5.

We also replaced the Frankot–Chellapa integrator by our multi-scale weighted Poisson integrator, described in a previous article [16]. Since the PhotoFace system is meant to operate automatically, we developed an algorithm to automatically extract the weight mask from the captured images. This algorithm is described in Section 6.

The new software takes about 3 s to obtain the normal map from the acquired photos, and another 3 s to compute the height map, on PhotoFace's PC. These times are less than 50% higher than those of the original PhotoFace software.

5. The EBPS algorithm

5.1. Table-based normal determination

In order to invert the shading signature function $\boldsymbol{l}$, we obtain a sufficiently dense set $\mathbb{T}$ of sample pairs $t_k = (\vec{n}_k, \boldsymbol{t}_k) \in \mathbb{S}^2 \times \mathbb{S}^{m-1}$ with $k = 1, 2 \ldots, N$, where $\boldsymbol{t}_k = \boldsymbol{l}(\vec{n}_k)$. Then for each pixel p in the target object we locate in this table the entry $(\vec{n}_r, \boldsymbol{t}_r)$ for which the distance $\|\boldsymbol{t}_r - \boldsymbol{s}[p]\|$ is minimum, and return the corresponding normal $\vec{n}_r$ as the presumed normal $\vec{s}[p]$ of the object's surface in that point.

This approach is very similar to that used by Woodham in 1994 [39]; except that we use normalized signatures $\boldsymbol{s}$, $\boldsymbol{l}$ instead of the unnormalized observation vectors. This approach is extremely flexible, since it can work with any light sources, concentrated or diffuse, and any constant isotropic finish β, as long as the lighting functions L_i are fairly smooth and the signature function $\boldsymbol{l}$ is invertible. Note that it does not require modeling the BRDF β or the light distributions Φ_i explicitly.

5.2. Fast table look-up

The accuracy of the result $\vec{n}_r$ returned by table look-up method depends only on the density of the sample normals $\vec{n}_k$ in $\mathbb{T}$ and the amount of noise present in the given images. As for the former, even if the normals in $\mathbb{T}$ are uniformly distributed over the hemisphere $\mathbb{H}^2$, the angular error between $\vec{n}_r$ and the true inverse $\boldsymbol{l}^{-1}(\boldsymbol{s}[p])$ will be about $1.5/\sqrt{N}$ radians. Thus, for example, in order to keep that part of the error below 1°, the table must have at least 8000 entries. For this reason, the table look-up step dominates the computational cost of photometric stereo.

Computing the distance $\|\boldsymbol{t}_k - \boldsymbol{s}[p]\|$ has cost proportional to m, therefore a simple linear search of the table would have cost proportional to Nm. Woodham's method to speed up the look-up was to quantize the observed radiances $S_i[p]$ as b-bit integers, and use them as indices into and m-dimensional array where the normals $\vec{n}_k$ were previously stored. This method reduced the look-up cost to $O(m)$; however, since the table required 2^{mb} entries, it severely limited the number of images m and the accuracy of the result. Other data structures for fast m-dimensional nearest neighbor searching have been proposed in the following years, such as k-dimensional tree search, Approximate nearest neighbors [29], or Locality Sensitive Hashing [43]. However, while these methods have $O(\log N)$ asymptotic look-up cost, they are not effective in this problem because of the so-called *curse of dimensionality* [44]: namely, the data becomes so sparse in high dimensional spaces that the methods only begin to work for very large tables, much larger than the tables needed for photometric stereo.

Therefore, we use instead a two-dimensional hashing method that we developed specifically for photometric stereo [45]. Our method exploits the fact that the set of all signatures $\boldsymbol{l}(\vec{n})$ for $\vec{n} \in \mathbb{H}^2$ is a two-dimensional surface patch in the positive orthant of $\mathbb{R}^m$. It uses a two-dimensional hashing array to reduce the search to a small number (constant, on the average) of table entries. Thus uses $O(N)$ space, and provides approximately $O(m)$ average look-up cost, independently of the table size N.

5.3. The reference objects

The reference object must be chosen so that its normals $\vec{g}[q]$ are accurately known and provide a dense and complete coverage of $\mathbb{H}^2$. Spherical or hemispherical reference objects with uniform albedo are most convenient, since they are easily obtained with highly accurate geometry, are completely described by a single geometric parameter (the radius), and allow $\vec{g}[q]$ to be computed directly from the coordinates of pixel q by simple algebraic formulas.

Since the example-based approach does not require explicit knowledge of the light source directions, we replaced the reflective

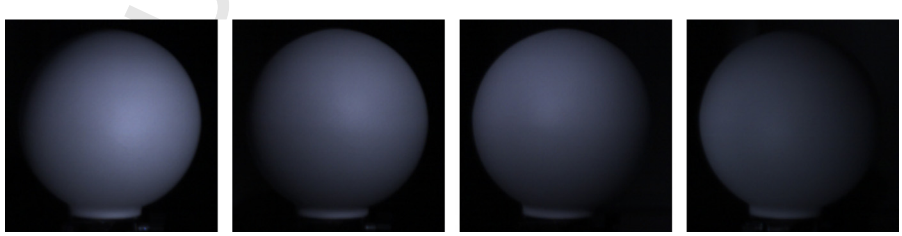


Fig. 6. Images of the example object captured by PhotoFace with near-infrared lighting, after ambient subtraction and cropping.

Please cite this article in press as: R.F.V. Saracchini, et al., Robust 3D face capture using example-based photometric stereo, Comput. Industry (2013), http://dx.doi.org/10.1016/j.compind.2013.04.003

sphere used in the original PhotoFace calibration run by the reference object, a sphere coated with semi-glossy white paint. See Fig. 6.

Although no effort was made to match the BRDF of this reference object to the BRDF of the human skin, the mere inclusion of a semi-glossy term improved considerably the accuracy of the computed normal maps; especially in the forehead and nose, where the glossy reflections were most conspicuous. See Section 7.

5.4. Using the example object images

Let $G_1, \ldots, G_m$ be the images of the reference object. Applying Eq. (1) to each pixel q in these images that falls on the reference object, we get

$$L_i(\vec{g}[q]) = \frac{G_i[q]}{\hat{g}[q]} \tag{7}$$

where $\vec{g}[q]$ and $\hat{g}[q]$ are the (known) surface normal and the albedo of the reference object at pixel q. Thus, for each pixel q on each image G_i of the reference object one obtains a sample value of the shading function L_i for a direction $\vec{g}[q]$. Therefore, every such pixel q provides a sample value of the shading signature function:

$$\boldsymbol{l}(\vec{g}[q]) = \frac{\mathbf{G}[q]}{\|\mathbf{G}[q]\|} \tag{8}$$

where

$$\mathbf{G}[q] = (G_1[q], G_2[q], \ldots, G_m[q]) \tag{9}$$

These samples of $\boldsymbol{l}$ are limited to the visible part of the object, and therefore to the directions $\vec{n} \in \mathbb{S}^2$ that make an acute angle with the viewing direction $\vec{v}$. We will denote by $\mathbb{H}^2$ that subset of $\mathbb{S}^2$.

5.5. Acquiring the shading functions

The signature table $\mathbb{T}$ could be built directly from the images of the reference object, namely $\vec{n}_k = \vec{g}[q]$ and $\boldsymbol{t}_k = \mathbf{g}[q]$ for each pixel q on the reference object. However these raw sample signatures are usually too few to yield the desired precision, and are often contaminated by imaging noise and by small geometrical defects (such as scratches and bumps) on the example object itself. Even if hardly perceptible on the images, these perturbations lead to large errors in the signature $\mathbf{g}[q]$ or in the normals $\vec{g}[q]$, thus introducing spurious data on the table. See Fig. 7.

In order to attenuate such errors and obtain a signature table that is dense enough, we fit a mathematical model $\tilde{L}_i$ of each shading function L_i to the raw data pairs $(\vec{g}[q], G_i[q]/\hat{g}[q])$. We then re-sample these mathematical models at a dense set of directions in $\mathbb{H}^2$ to obtain the table $\mathbb{T}$.

5.6. Shading function model

We now describe how to obtain the approximations $\tilde{L}_i$. Since this step is carried out separately for each image, we will drop the index i from L_i, $\tilde{L}_i$ and G_i in this section.

We use a linear approximation model with weighted least squares criterion. Specifically, we choose a set of *basis functions* $\phi_1, \phi_2 \ldots \phi_l : \mathbb{S}^2 \Rightarrow \mathbb{R}$, and look for an approximation $\tilde{L}$ of L of the form

$$\tilde{L}(\vec{n}) = \sum_{r=1}^{l} \alpha_r \phi_r(\vec{n}) \tag{10}$$

for each $\vec{n} \in \mathbb{S}^2$, where $\alpha_1, \ldots, \alpha_l$ are real coefficients chosen to minimize the weighted quadratic error $Q(\tilde{L})$, namely

$$Q(\tilde{L}) = \sum_q \left(\frac{G[q]}{\hat{g}[q]} - \tilde{L}(\vec{g}[q]) \right)^2 \tag{11}$$

The sum here is taken over all pixels q that are completely inside the outline of the reference object. The albedo $\hat{g}[q]$ must be given, and the coefficient vector α can be found by solving the linear system

$$M\alpha = b \tag{12}$$

where M is a $l \times l$ matrix and b is a column vector of l elements given by

$$\begin{aligned} M_{ij} &= \sum_q \phi_i(\vec{g}[q]) \phi_j(\vec{g}[q]) \\ b_i &= \sum_q \phi_i(\vec{g}[q]) G[q]/\hat{g}[q] \end{aligned} \tag{13}$$

The basis we have chosen for this application consists of the monomials $\phi_i(\vec{n}) = x^r y^s z^t$ for all natural numbers r, s, t whose sum is either d or $d-1$, where d is a chosen positive integer, in some arbitrary order. Here x, y, z are the Cartesian coordinates of the normal vector $\vec{n}$. These monomials generate precisely the space of all spherical harmonic functions of maximum degree d [46], but they are much easier to compute than the standard spherical harmonic basis functions Y_{rs} [47]. The latter are usually preferred because they are orthogonal when integrated over the whole sphere. However, the inner product implicitly used in the quadratic error formula (11) is a discrete sums over an irregular set of normals $\vec{g}[q] \in \mathbb{H}^2$. The harmonic basis functions Y_{rs} are not orthogonal in this inner product, and therefore have no clear advantage over the monomials.

5.7. Virtual reference objects

For checking purposes, the fitted shading functions $\tilde{L}_i$ an be used to produce synthetic images $\tilde{G}_i$ of a “virtual” reference object,

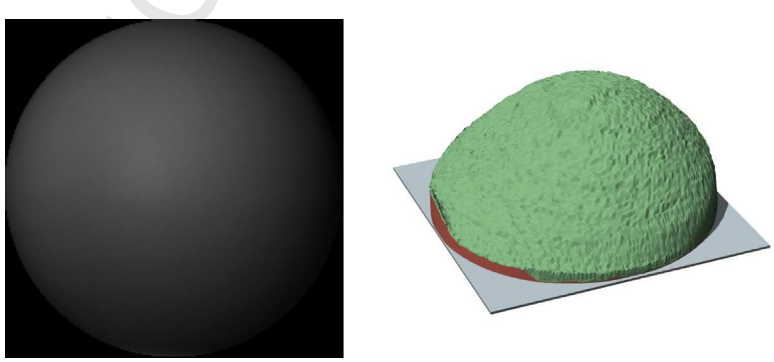


Fig. 7. An image G_i of a reference object and a plot of its shading function $L_i(\vec{n})$ for $\vec{n} \in \mathbb{H}^2$ hemisphere. Note the substantial amount of noise in the latter that is not visible in the former.

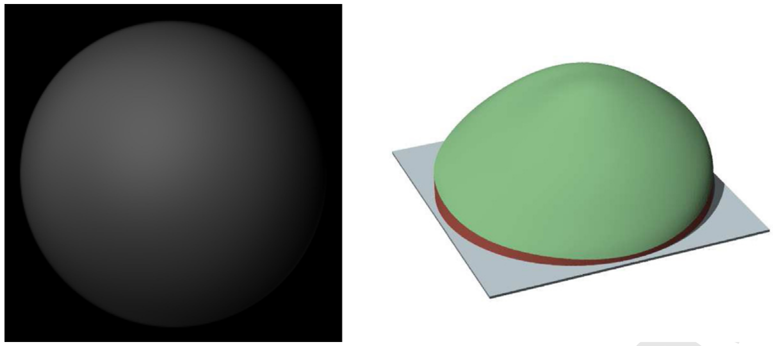


Fig. 8. Synthetic image of a virtual reference object (left) using a smoothed shading function $\tilde{L}_i$ (right) fitted to the data of the real reference object image of Fig. 7.

where each pixel $\tilde{G}_i[q]$ is painted with the color $\tilde{L}_i(\vec{g}[q])$ expected from its assumed normal $\vec{g}[q]$. These new synthetic images can be directly compared to the raw images $G[i]$ of the reference object. See Fig. 8

6. Obtaining the weight map

We now consider the problem of obtaining the gradient reliability weight map that is required by the robust integrators. There are several published methods that try to derive the weights from the gradient map itself, by using the fact that the gradient of a function must be a curl-free vector field. Namely, the *Y*-derivative of the *X*-slope must be equal to the *X*-derivative of the *Y*-slope; the difference between the two being the curl of the field. If the curl is not zero, the gradient map cannot be integrated. Those methods generally mark as unreliable those pixels where the curl is not zero, or where the gradient of the integrated map does not match the given gradient [34,35].

However, the zero-curl condition is only necessary, but not sufficient, for the gradient to be correct. In Fig. 2, the curl is non-zero along the sides of the ramp, but is zero everywhere else. As that example shows, the discontinuities cannot be detected from the gradient map alone. In general, the weight map must be determined by problem-specific methods, for example edge detection by projected shadows [48].

6.1. A masking algorithm for PhotoFace

We have developed an algorithm for automatic extraction of the reliability weight mask from the PhotoFace image sets. The method aims to exclude the areas of the scene where photometric stereo cannot be applied, such as the distant background, hair, eye pupils, and areas which are illuminated by only two of the four sources (right below the nose and chin, and on the temples). The method also uses specific heuristics to exclude clothing and other disconnected bits of surface, such as parts of the ear.

Fig. 9. Construction of the weight mask for a captured image set. Top left: the initial weights $w[p]$ computed from the signature match quality $\|\mathbf{s}[p] - \mathbf{t}_r\|$ and from the computed albedo $\hat{s}[p]$. Top center: the pixels that pass the slope test $\vec{s}[p]\cdot\vec{v} > \varepsilon_1$. Top right: the weight mask after removing pixels with weight below ε_2. Bottom left: after morphological opening. Bottom center: after removing the disconnected parts. Bottom right: after morphological closing.

Please cite this article in press as: R.F.V. Saracchini, et al., Robust 3D face capture using example-based photometric stereo, Comput. Industry (2013), http://dx.doi.org/10.1016/j.compind.2013.04.003

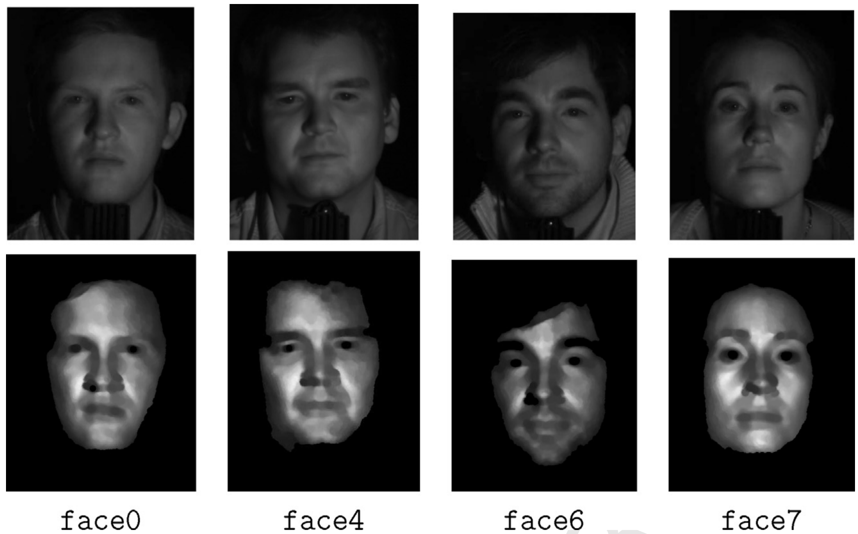


Fig. 10. Top row: image S_0 from each of the four test datasets captured with PhotoFace, with near-infrared lighting, after ambient light subtraction. Bottom row: the corresponding weight masks for integration.

There are several published algorithms for the detection of human skin in photographs. [49,50]. However, those algorithms generally rely on color information which is not available in the near-infrared monochromatic images captured by PhotoFace. Instead, we use four main criteria: the self-consistency of the photometric stereo computation, the estimated albedo of the surface, the angle between the computed normal and the viewing direction, spatial coherence, and continuity.

Specifically, we first set the weigh $w[p]$ by the formula

$$w[p] = \|\boldsymbol{S}[p]\|^2 e^{-1/2d^2} \tag{14}$$

where

$$d = \frac{\|\boldsymbol{s}[p] - \boldsymbol{t}_r\|\|\boldsymbol{S}[p]\|}{\sigma} \tag{15}$$

and $\boldsymbol{t}_r$ is the signature from the table $\mathbb{T}$ that best matches the pixel's signature $\boldsymbol{s}[p]$. The first factor eliminates in formula (14) dark areas where the signature $\boldsymbol{s}[p]$ is too contaminated by noise. The second factor penalizes pixels where the observed signature $\boldsymbol{s}[p]$ deviates from the expected signature for the recovered normal, and is therefore likely to be contaminated by noise, cast shadows, or other un-modeled effects.

Next, we set $w[p]$ to zero if the Z component $\vec{s}[p]\cdot\vec{v}$ of the computed normal $\vec{s}[p]$ is less than a fixed threshold ε_1 (currently set to 0.03). This step eliminates parts where the surface seems to be nearly perpendicular to the viewing direction, and therefore cannot be simultaneously illuminated by three light sources. Then the weight $w[p]$ is set to zero if it is less than another threshold ε_2 (currently set to 0.2), in order to completely remove from the computation those pixels whose height is expected to be too unreliable to use.

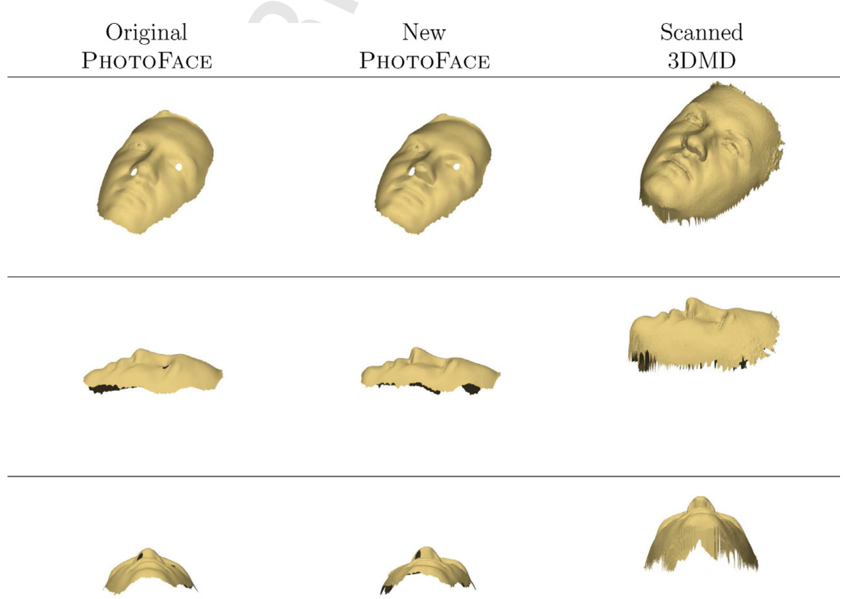


Fig. 11. Height maps computed for the test dataset `face0`. The height maps were rendered as 3D surfaces with arbitrary illumination, from three viewpoints: oblique (top row), profile (middle row), and from below (bottom row).

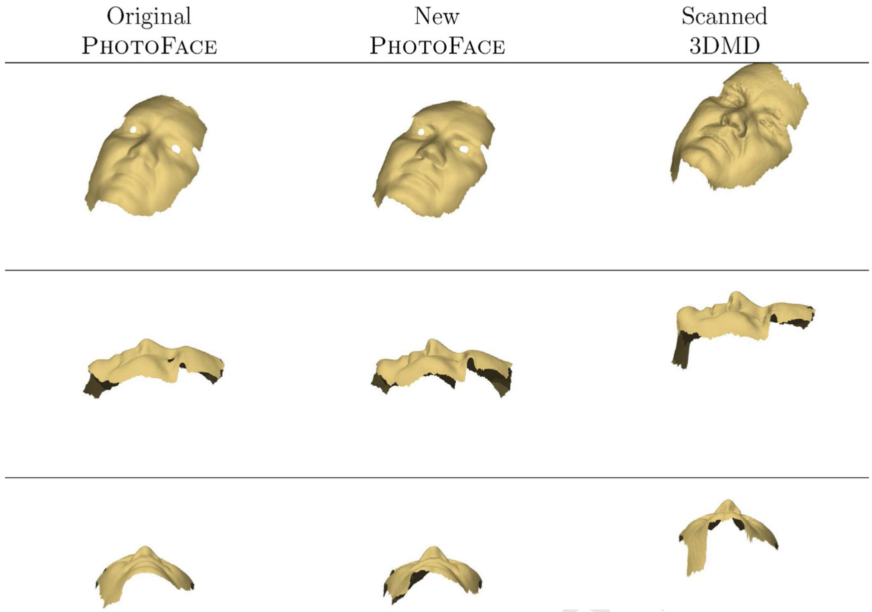


Fig. 12. Height maps computed for the test dataset `face4`.

Next, we apply a gray-scale morphological opening operation (an erosion followed by a dilatation) with a 5×5 circular kernel, to remove small isolated pixel clumps and break any narrow "bridges", as well as removing pixels near problematic areas such as nostrils and eyes. Then we identify the connected components of the image (separated by areas of zero weight), and discard all but the most central one (which is assumed to be the face, considering how the image was cropped). Finally, we apply a gray-scale morphological closing operation (a dilation followed by an erosion) with a 6×6 circular kernel, meant to preserve the "holes" in the mask at the eye pupils, but close other small holes and gaps. See Fig. 9.

7. Tests

To illustrate the changes, we show below the height maps obtained with the old and new PhotoFace algorithms, on the same four datasets labeled `face0`,`face4`,`face6`,`face7`. See Fig. 10.

The normals were computed from these datasets with the old PhotoFace analytic algorithm and with our new EBPS algorithm. For the latter, shading functions were fitted to the four images of the reference object (Fig. 6) using the monomial basis with maximum degree 6. These fitted functions were then re-sampled at approximately 10,000 evenly distributed normal directions in $\mathbb{H}^2$ to form the signature table $\mathbb{T}$.

Both normal maps, old and new, were then integrated with our multi-scale Poisson-based integrator. The resulting height maps are shown in Figs. 11–14. For comparison, we also show the facial geometries of those same four people, captured in separate occasions by the 3DMD face scanner [6] and cropped to about the same part of the face. (Unfortunately, numerical comparison with the latter is not viable due to differences in facial expression and perspective distortion.)

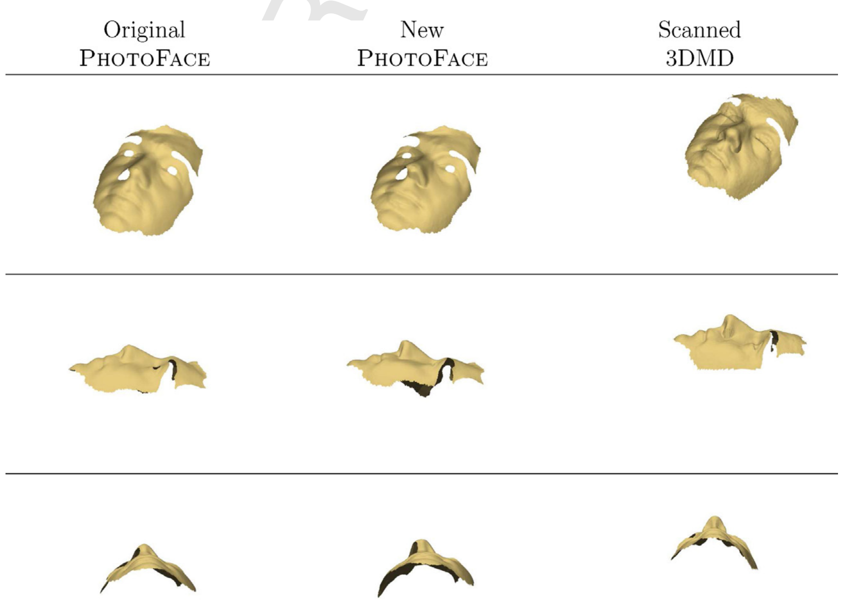


Fig. 13. Height maps computed for the test dataset `face6`.

Please cite this article in press as: R.F.V. Saracchini, et al., Robust 3D face capture using example-based photometric stereo, Comput. Industry (2013), http://dx.doi.org/10.1016/j.compind.2013.04.003

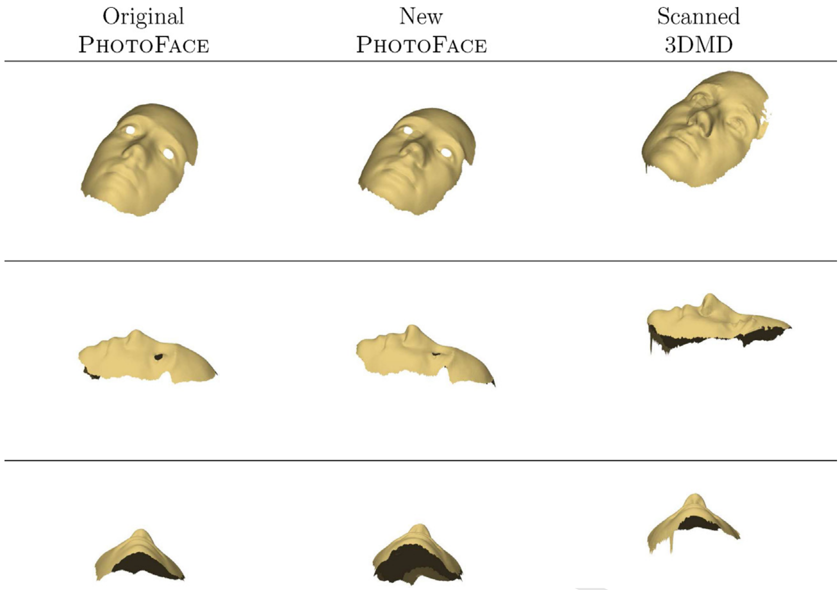


Fig. 14. Height maps computed for the test dataset `face7`.

8. Conclusions

We showed how example-based photometric stereo is a viable alternative for the capture the 3D surface geometry of human faces, with unique advantages including low cost, high speed, non-intrusiveness, flexibility, high resolution, large working distance, and indifference to albedo and ambient lighting. We also pointed out the inadequacy of the popular Frankot–Chellapa Fourier-based gradient map integrator compared to Poisson-based integrators. We described an algorithm for automatic generation of the weight mask needed by those integrators.

In particular, we showed that the accuracy and robustness of PhotoFace, a state-of-the art photometric face capture system, are significantly improved, with little extra computation cost, when the of popular analytic algorithms for normal computation are replaced by EBPS, and the Frankot–Chellapa integrator is replaced by our multi-scale integrator et al. [16]. Even though no effort was made to reproduce the BRDF of the human skin in the reference object, the mere inclusion of a glossy term (which cannot be handled by analytic methods) was enough to remove most of the distortions created by glossy reflections when using the old algorithms.

Even with our improvements, the height maps obtained with current version PhotoFace have noticeable problems, especially in the region of nostrils and under the chin. These defects are not due to algorithm limitations but rather to insufficient illumination in those areas. The addition of two more light sources to the prototype would suffice to ensure the basic requirement of photometric stereo, namely that every point of the target surface be illuminated by at least three light sources.

References

[1] W. Zhao, R. Chellappa, P.J. Phillips, A. Rosenfeld, Face recognition: A literature survey, ACM Computing Surveys 35 (2003) 399–458. , http://dx.doi.org/10.1145/954339.954342.

[2] K.W. Bowyer, K. Chang, P. Flynn, A survey of approaches and challenges in 3d and multi-modal 3d + 2d face recognition, Computer Vision and Image Understanding 101 (1) (2006) 1–15. , http://dx.doi.org/10.1016/j.cviu.2005.05.005 http://www.sciencedirect.com/science/article/pii/S1077314205000822.

[3] S. Berretti, A. Del Bimbo, P. Pala, 3d face recognition using isogeodesic stripes, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (12) (2010) 2162–2177.

[4] Y. Wang, J. Liu, X. Tang, Robust 3d face recognition by local shape difference boosting, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (10) (2010) 1858–1870.

[5] Konica Minolta Sensing, 2012, http://sensing.konicaminolta.us (accessed 22.08.12).

[6] 3dmdface system, 2010, http://www.3dmd.com/3dmdface.html (accessed 22.04.10).

[7] Dimensional imaging's di4d 4d facial capture system, 2013, http://www.di3d.com/ (accessed 03.03.13).

[8] W. Smith, E. Hancock, Facial shape-from-shading and recognition using principal geodesic analysis and robust statistics, International Journal of Computer Vision 76 (2008) 71–91.

[9] V. Blanz, T. Vetter, A morphable model for the synthesis of 3d faces, in: Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH'99, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 1999, pp. 187–194. , http://dx.doi.org/10.1145/311535.311556.

[10] M.F. Hansen, G.A. Atkinson, L.N. Smith, M.L. Smith, 3d face reconstructions from photometric stereo using near infrared and visible light, Computer Vision and Image Understanding (2010) in press.

[11] V. Nozick, Pyramidal normal map integration for real-time photometric stereo, EAM Mechatronics 2010 (2010) 128–132.

[12] G. Vogiatzis, C. Hernández, Self-calibrated, multi-spectral photometric stereo for 3d face capture, International Journal of Computer Vision (2011) 1–13.

[13] L. Broadbent, K. Emrith, A.R. Farooq, M.L. Smith, L.N. Smith, 2.5 d facial expression recognition using photometric stereo and the area weighted histogram of shape index, in: IEEE RO-MAN, 2012, 2012, 490–495.

[14] L. Broadbent, K. Emrith, A. Farooq, M. Smith, L. Smith, 2.5d facial expression recognition using photometric stereo and the area weighted histogram of shape index, in: IEEE RO-MAN, 2012, 2012, 490–495. , http://dx.doi.org/10.1109/RO-MAN.2012.6343799.

[15] S.N. Kautkar, G.A. Atkinson, M.L. Smith, Face recognition in 2d and 2.5d using ridgelets and photometric stereo, Pattern Recognition.

[16] R.F.V. Saracchini, J. Stolfi, H.C. Leitão, G.A. Atkinson, M.L. Smith, A robust multi-scale integration method to obtain the depth from gradient maps, Computer Vision and Image Understanding 116 (8) (2012) 882–895.

[17] R.J. Woodham, Determining surface curvature with photometric stereo, in: Proceedings of the 1989 IEEE International Conference on Robotics and Automation, vol. 1, 1989, pp. 36–42.

[18] S. Barsky, M. Petrou, The 4-source photometric stereo technique for three-dimensional surfaces in presence of highlights and shadows, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (10) (2003) 1239–1252.

[19] C. Yu, Y. Seo, S. Lee, Photometric stereo from maximum feasible lambertian reflections, in: K. Daniilidis, P. Maragos, N. Paragios (Eds.), Computer Vision ECCV 2010, vol. 6314 of Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2010, pp. 115–126.

[20] H. Hayakawa, Photometric stereo under a light source with arbitrary motion, Journal of the Optical Society of America Series A (JOSAA) 11 (11) (1994) 3079–3089.

[21] A. Yuille, D. Snow, Shape and albedo from multiple images using integrability, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 97), 1999, pp. 158–164.

[22] A. Yuille, D. Snow, R. Epstein, P. Belhumeur, Determining generative models of objects under varying illumination: Shape and albedo from multiple images using

Please cite this article in press as: R.F.V. Saracchini, et al., Robust 3D face capture using example-based photometric stereo, Comput. Industry (2013), http://dx.doi.org/10.1016/j.compind.2013.04.003

svd and integrability, International Journal of Computer Vision 35 (1999) 203–222. , http://dx.doi.org/10.1023/A:1008180726317.

[23] R. Basri, D. Jacobs, Lambertian reflectance and linear subspaces, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (2) (2003) 218–233.

[24] R. Basri, D. Jacobs, I. Kemelmacher, Photometric stereo with general unknown lighting, International Journal of Computer Vision 72 (3) (2007) 239–257.

[25] N. Alldrin, T. Zickler, D. Kriegman, Photometric stereo with non-parametric and spatially-varying reflectance, in: IEEE Conference on Computer Vision and Pattern Recognition, 2008, CVPR 2008, (2008), pp. 1–8. , http://dx.doi.org/10.1109/CVPR.2008.4587656.

[26] D. Goldman, B. Curless, A. Hertzmann, S. Seitz, Shape and spatially-varying brdfs from photometric stereo, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (6) (2010) 1060–1071. , http://dx.doi.org/10.1109/TPAMI.2009.102.

[27] G. McGunnigle, Photometric stereo with gradated extended sources for recovery of specular surfaces, Journal of the Optical Society of America A 27 (5) (2010) 1127–1136. , http://dx.doi.org/10.1364/JOSAA.27.001127 http://josaa.osa.org/abstract.cfm?URI=josaa-27-5-1127.

[28] T. Higo, Y. Matsushita, K. Ikeuchi, Consensus photometric stereo, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010, 1157–1164.

[29] A. Hertzmann, S.M. Seitz, Example-based photometric stereo: shape reconstruction with general varying BRDFs, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (8) (2005) 1254–1264.

[30] A. Hertzmann, S.M. Seitz, Shape and materials by example: a photometric stereo approach, in: Proceedings IEEE CVPR 2003, vol. 1, 2003, pp. 533–540. http://citeseer.ist.psu.edu/article/hertzmann03shape.html.

[31] H.C.G. Leitão, R.F.V. Saracchini, J. Stolfi, Matching photometric observation vectors with shadows and variable albedo, in: Proceedings of the 21st Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI), 2008, pp. 179–186.

[32] R. Fraile, E.R. Hancock, Combinatorial surface integration, in: Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06), vol. 1, 2006, pp. 59–62.

[33] R.T. Frankot, R. Chellappa, A method for enforcing integrability in shape from shading algorithms, IEEE Transactions on Pattern Analysis and Machine Intelligence 10 (4) (1988) 439–451.

[34] A. Agrawal, R. Raskar, R. Chellappa, What is the range of surface reconstructions from a gradient field? in: Proceedings of the 9th European Conference on Computer Vision (ECCV), vol. 3951, 2006, pp. 578–591.

[35] D. Reddy, A. Agrawal, R. Chellappa, Enforcing integrability by error correction using ℓ_1-minimization, in: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 2350–2357.

[36] R.F.V. Saracchini, J. Stolfi, H.C. da Gama Leitão, G.A. Atkinson, M.L. Smith, Multiscale integration of slope data on an irregular mesh, in: PSIVT (1), 2011, 109–120.

[37] A. Georghiades, P. Belhumeur, D. Kriegman, From few to many: illumination cone models for face recognition under variable lighting and pose, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (6) (2001) 643–660.

[38] S.-W. Lee, P.S. Wang, N.Y. Svetlana, L. Seong-whan, Noniterative 3d face reconstruction based on photometric stereo, International Journal of Pattern Recognition and Artificial Intelligence 22 (03) (2008) 389–410.

[39] R.J. Woodham, Gradient and curvature from the photometric stereo method including local confidence estimation, Journal of the Optical Society of America, Series A 11 (11) (1994) 3050–3068.

[40] G. Schindler, Photometric stereo via computer screen lighting, in: Proceedings of the 4th International Symposium on 3D Data Processing, Visualization and Transmission, 2008.

[41] G. Fyffe, X. Yu, P. Debevec, Single-shot photometric stereo by spectral multiplexing, in: IEEE International Conference on Computational Photography (ICCP), 2011, 2011, 1–6. , http://dx.doi.org/10.1109/ICCPHOT.2011.5753116.

[42] R. Lienhart, J. Maydt, An extended set of haar-like features for rapid object detection, in: Proceedings of the 2002 International Conference on Image Processing, vol. 1, 2002, pp. I-900–I-903. , http://dx.doi.org/10.1109/ICIP.2002.1038171.

[43] L. Zhong, J.J. Little, Photometric stereo via locality sensitive high-dimension hashing, in: in: Proceedings of the Second Canadian Conf. on Computer and Robot Vision (CRV'05), 2005, pp. 104–111.

[44] P. Indyk, R. Motwani, Approximate nearest neighbour: toward removing the curse of dimensionality, in: Proceedings of the 13th Annual ACM Symposium on Theory of Computing (STOC'98), 1998, pp. 604–613.

[45] R.F.V. Saracchini, J. Stolfi, H.C. da Gama Leitao, A uniform grid structure to speed up example-based photometric stereo, IEEE Transactions on Image Processing 20 (12) (2011) 3495–3507. , http://dx.doi.org/10.1109/TIP.2011.2159386.

[46] A. Gomide, J. Stolfi, Bases for non-homogeneous polynomial $\mathbf{C}_k$ splines on the sphere, in: Proceedings of LATIN'98 - Latin American Theoretical Informatics Conference, vol. 1380 of Lecture Notes in Computer Science, Springer, 1998, pp. 133–140.

[47] C. Müller, Spherical Harmonics, vol. 17, Springer, 1966.

[48] M. Chandraker, S. Agarwal, D. Kriegman, ShadowCuts: photometric stereo with shadows, in: CVPR07, 2007, 1–8.

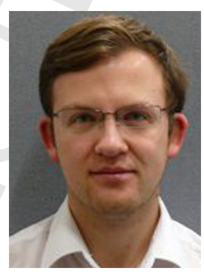
[49] C. Garcia, G. Tziritа, Face detection using quantized skin color regions merging and wavelet packet analysis, in: IEEE Transactions on Multimedia, vol. 3, 1999, 264–277.

[50] R. lien Hsu, M. Abdel-mottaleb, A.K. Jain, Face detection in color images, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (2002) 696–706.

Rafael F.V. Saracchini received the B.Sc. degree from the Fluminense Federal University, Niterói, Brazil, in 2007, and the Ph.D. degree from State University of Campinas, Campinas, Brazil, in 2012. His research interest include 3-D data recovery through digital images using geometric and photometric stereo from non-Lambertian and glossy/shiny surfaces and its applications in medicine, industry, and security. In 2012 Rafael moved to the Technical Institute of Castilla y León to develop Computer Vision aided systems focused in assisted living to elderly people and 3D reconstruction of indoors environments.

Jorge Stolfi received the B.E. degree in electrical engineering and the M.Sc. degree in computer science from the University of Sao Paolo, Sao Paolo, Brazil, in 1973 and 1979, respectively, and the Ph.D. degree in computer science from Stanford University, Stanford, CA, in 1989. He is currently a Full Professor with the Institute of Computing, State University of Campinas, Campinas, Brazil. His research interests include natural language processing, computational geometry, computer graphics, numerical analysis, image processing, and applications.

Helena Cristina da Gama Leitão received the B.Sc. and M.Sc. degrees in computer science from the Federal University of Rio de Janeiro, Rio de Janeiro, Brazil, in 1988 and 1993, respectively, and the Ph.D. degree from the State University of Campinas, Campinas, Brazil, in 1999. She is currently an associate Professor with the Institute of Computing, Fluminense Federal University, Niterói, Brazil. Her research interests include pattern recognition for archeological, medical, and molecular biology applications.



Gary Atkinson completed an M.Sc. degree in physics at the University of Nottingham in 2003. Upon graduation, he moved to the University of York to study for a Ph.D. degree in the Department of Computer Science, under the supervision of Edwin Hancock. His research was concerned with improving shape recovery algorithms and reflectance function estimation for computer vision. Most of his work involved the exploitation of the polarizing properties of reflection from surfaces. In 2007, Gary moved to UWE to work on face reconstruction and recognition research in collaboration with Imperial College London. Gary also has ties with the University of Campinas, the University of York, the University of Central Lancashire and several industrial partners. He supervised Mark Hansen's PhD in face recognition to completion in March 2012. In addition to his computer vision research, for which he has published over ten major peer-reviewed international journal papers, Gary has worked in the medical field to develop a new means of assessing head deformations in babies affected by plagiocephaly. In 2010, Gary acted as the editor and UK and Europe co-convener for the International Conference and Exhibition on Biometrics Technology in Coimbatore, India. Gary teaches undergraduate and postgraduate courses in Computer Vision and Engineering."

Melvyn Smith is Professor of Machine Vision and Director of the Centre for Machine Vision in the Bristol Robotics Laboratory (BRL) a partnership between the University of Bristol and the University of the West of England. He received his B.Eng. (Hons) degree in Mechanical Engineering from the University of Bath in 1987, M.Sc. in Robotics and Advanced Manufacturing Systems from the Cranfield Institute of Technology in 1988 and Ph.D. in Computer Vision and Graphical Modeling from the UWE in 1997. He acts as associate editor for four leading international journals, including Computers in Industry, and regularly acts as a program committee member for international conferences in the computer vision field. To date he has published in excess of 150 articles, including 4 patent applications, a book and 2 book chapters, edited 2 journal special issues, 42 refereed journal papers, 59 refereed conference papers as well as numerous non-refereed papers and special reports. He regularly presents his work through invited seminars, as well as articles and interviews to the popular media. He has been a member of the EPSRC Peer Review College since 2003, served as a European commission candidate evaluator for the Sixth Framework and is currently a program and evaluator/review expert and monitoring expert for the EU Framework 7 Programme. Prof. Smith is a Chartered Engineer and an active member of the IET.