Color Photometric Stereo for Multicolored Surfaces

Robert Anderson
University of Cambridge
ra312@cam.ac.uk

Björn Stenger
Toshiba Research Europe
bjorn.stenger@crl.toshiba.co.uk

Roberto Cipolla
University of Cambridge
cipolla@eng.cam.ac.uk

Abstract

We present a multispectral photometric stereo method for capturing geometry of deforming surfaces. A novel photometric calibration technique allows calibration of scenes containing multiple piecewise constant chromaticities. This method estimates per-pixel photometric properties, then uses a RANSAC-based approach to estimate the dominant chromaticities in the scene. A likelihood term is developed linking surface normal, image intensity and photometric properties, which allows estimating the number of chromaticities present in a scene to be framed as a model estimation problem. The Bayesian Information Criterion is applied to automatically estimate the number of chromaticities present during calibration. A two-camera stereo system provides low resolution geometry, allowing the likelihood term to be used in segmenting new images into regions of constant chromaticity. This segmentation is carried out in a Markov Random Field framework and allows the correct photometric properties to be used at each pixel to estimate a dense normal map. Results are shown on several challenging real-world sequences, demonstrating state-of-the-art results using only two cameras and three light sources. Quantitative evaluation is provided against synthetic ground truth data.

1. Introduction

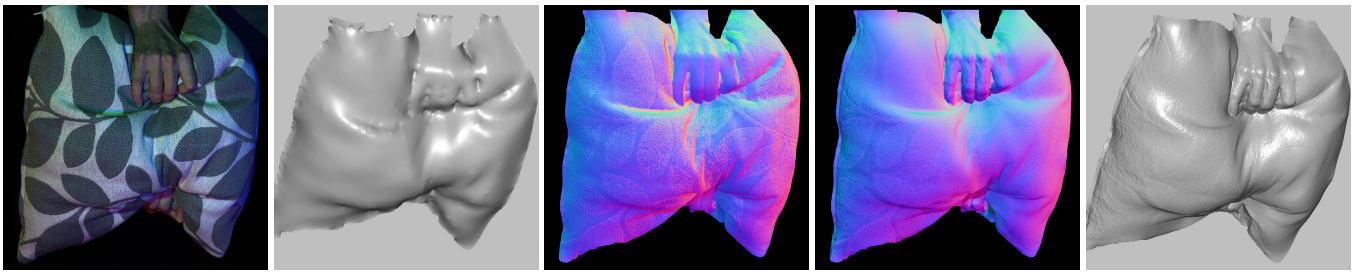
Capture of deforming surfaces is becoming increasingly important for a variety of applications in graphics, medical imaging, and analysis of deployable structures. Practical methods of acquiring high resolution geometry in both the spatial and the temporal domains are required. In this paper we propose a system that is capable of this and has only modest equipment and computational requirements.

The proposed approach is based upon color (or multispectral) photometric stereo which uses three different colored lights to capture three lighting directions in a single RGB image. This allows photometric stereo to be carried out on each frame of a video sequence, generating high resolution geometry at the same frame rate as the input video. One major weakness of standard multispectral photometric stereo is its assumption that the observed scene is of constant chromaticity. Generalization to scenes with varying chromaticity has only been dealt with before by either resorting to time multiplexing [7, 16], to regularization of the normal field [14] or to the use of a depth camera to provide extra information [3]. In this paper we propose a similar approach to that of [3] but using a more principled framework and without the need of a depth camera.

1.1. Prior work

Multispectral photometric stereo was first demonstrated over 15 years ago by several groups [8, 19, 26]. More recently it has been used in a variety of applications [6, 11, 15, 17]. Johnson and Adelson [15] simplify calibration by imaging an indenter with known reflective properties but this is not applicable to general scene capture. Calibration is carried out by Klaudiny *et al*. [17] first by estimating lighting directions using a specular sphere and then cycling one of the light sources through each of the three colors to estimate an average calibration for the whole scene. A more robust approach is proposed by Hernandez and Vogiatzis in [11] in which a sequence of rigid body motions of the target object allows approximate geometry to be reconstructed using structure from motion. This provides a set of image normal pairs which allow for estimation of lighting directions and surface reflectance.

The assumption of constant chromaticity made by multispectral photometric stereo can be relaxed by the addition of time multiplexing, as shown by DeDecker *et al*. [7] and Kim *et al*. [16]. These systems allow for reconstruction of scenes with varying chromaticities but at the expense of requiring at least two frames of input to produce one frame of geometry, halving the temporal resolution of their results. Janko *et al*. [14] avoid the need for time multiplexing by tracking texture on the surface and optimizing both surface chromaticity and normal direction over a complete sequence. To make this tractable, regularization is required on the normal field which can smooth over the fine detail that photometric stereo is otherwise capable of capturing. Current state-of-the-art photometric reconstructions are achieved through



(a) (b) (c) (d) (e)

Figure 1. **Reconstruction overview**. (a) Input image. (b) Low-resolution geometry recovered from two-view stereo. (c) The normal map estimated using a constant chromaticity assumption contains large errors. (d) The normal map estimated using the proposed approach removes these errors. (e) Final high quality geometry obtained by combining stereo reconstruction with normal map.

purely time multiplexing. Ma *et al*. [21] use a combination of structured light and time-multiplexed spherical illumination patterns to achieve high quality results at the expense of a complex equipment setup.

Other methods for dynamic geometry capture produce results with very different characteristics. Multiview stereo is a well studied field with current systems, such as Furukawa and Ponce's PMVS software [10], capable of reconstructing a wide variety of scenes. Faces and other objects which exhibit little texture have traditionally been challenging to reconstruct using multiview stereo, but recently both Bradley *et al*. [5] and Beeler *et al*. [4] have demonstrated accurate results. Reconstruction can be facilitated by adding texture to the target surface by either projecting light patterns, [27], or by applying makeup, [9]. Phase shift structured light gives similar quality results as demonstrated by Weise *et al*. [25].

Whilst multiview stereo and structured light systems are capable of providing accurate global geometry they are generally limited in the resolution of features they are able to reconstruct. The idea of using other cues which contain high frequency information to augment the output from multiview stereo is not new, for example Ikeuchi [13] demonstrated the fusion of multiview and photometric stereo in 1987. More recently Beeler *et al*. [4] used a qualitative method based on shading to improve their reconstructions. Several systems [1, 2, 17, 21] have demonstrated the effectiveness of using photometric stereo to provide accurate high frequency information with which to complement a low frequency reconstruction obtained by stereo or structured light. One recent example is the work of Vlasic *et al*. [24], showing detailed capture of high resolution , watertight models of human bodies. The light sphere setup, consisting of several hundred controllable LEDs, used in this work is however prohibitively complex for some applications, and the time-multiplexed photometric stereo requires at least three input frames for each reconstruction.

1.2. Contributions

The theoretical contributions of this paper are firstly to present a novel calibration technique for multispectral photometric stereo that can be applied to objects with multiple piecewise constant chromaticities. Secondly we demonstrate how to automatically estimate the number of chromaticities present during this calibration. Furthermore we show how given an approximate normal map we can estimate the correct photometric properties to use at each pixel of a new image allowing for dense normal map extraction.

The practical contribution of this paper is to demonstrate that state-of-the-art results can be achieved with a modest hardware setup consisting of two cameras and three passive light sources. Time multiplexing is avoided and no markers are added to the target. Qualitative results are demonstrated on challenging real world sequences and a quantitative analysis is carried out on synthetic data.

2. System overview

The proposed system uses two complementary approaches to capture the geometry of deforming surfaces. Two cameras allow multiview stereo to reconstruct a low-resolution depth map. This is then augmented using normals obtained from multispectral photometric stereo.

The stereo reconstruction uses the algorithm of Beeler *et al*. [4]. The combination of low frequency depth information and high frequency normal information is carried out using the technique of Nehab *et al*. [22]. The novelty of the system lies in the photometric reconstruction as detailed in the following sections.

3. Photometric reconstruction

3.1. Multiple chromaticities

Given an image of a surface of constant chromaticity, illuminated by three spectrally and spatially separated light sources, it is well known that it is possible to estimate a surface normal at each pixel [8]. Given a surface with $N > 1$

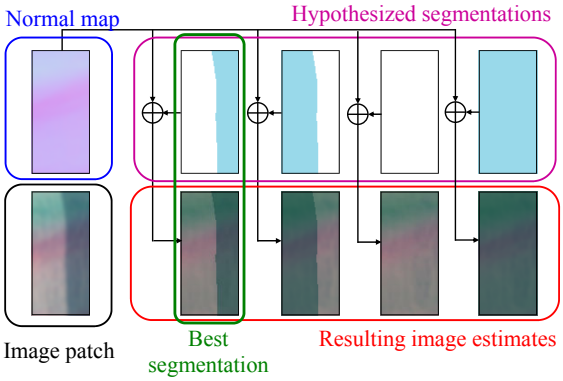


Figure 2. **Evaluating segmentation hypotheses.** Given a normal map (blue) and a hypothesized segmentation (magenta) an image estimate can be formed (red). The segmentation producing the estimate that best matches the input image (black) is chosen.

regions of different chromaticity this is no longer possible as a change in pixel color could be caused by either a change in surface normal or a change in surface chromaticity. To resolve this ambiguity an input image can be segmented into regions of constant chromaticity before normal estimation is carried out. To perform this segmentation we use the low-resolution stereo depth map to compute a smoothed normal direction at each pixel. For each of the N chromaticities in the scene the smoothed normal at any pixel will predict a different color as shown in figure 2 for the simple case of $N = 2$. A good segmentation can be found by ensuring that an image generated from the smoothed normals matches the observed image as closely as possible. In carrying out a segmentation we wish to enforce two constraints:

1. The likelihood of generating the observed image from the smoothed normal map is maximized.

2. Chromaticity is locally constant.

In section 3.2 we develop an expression for the likelihood term necessary to enforce the first constraint and in section 3.3 we show how a Markov Random Field (MRF) can be used to carry out the segmentation whilst enforcing the second constraint.

3.2. Likelihood term

We assume a Lambertian reflectance model and ensure that there is no ambient lighting. Under these conditions, given three distant point light sources illuminating a surface with unit normal $\mathbf{n}$ and albedo α, it has been shown, [8], that the observed intensity of the surface is given by

$$\mathbf{c} = \alpha \mathbf{V}\mathbf{L}\mathbf{n} = \begin{bmatrix} \mathbf{v_0} & \mathbf{v_1} & \mathbf{v_2} \end{bmatrix} \begin{bmatrix} \mathbf{l_0} & \mathbf{l_1} & \mathbf{l_2} \end{bmatrix}^\top \alpha \mathbf{n}, \quad (1)$$

where $\mathbf{c}$, $\mathbf{l}_i$ and $\mathbf{v}_i$ are all column vectors of length three. $\mathbf{c}$ denotes the RGB image intensity, $\mathbf{l}_i$ defines the direction of light i and $\mathbf{v}_i$ is the combined response of surface and sensor to light i. The matrix $\mathbf{V}$ models the combination of the surface's chromaticity, the lights' spectral distributions and the camera sensors' spectral sensitivities. It is this matrix that varies for regions of different chromaticity. The albedo of the surface α is a value between zero and one which is equal to the proportion of incoming light reflected by the surface.

We assume that each channel of the image is corrupted by additive white Gaussian noise with variance σ^2 at each pixel independently, making $\mathbf{c}$ normally distributed with

$$p\left(\mathbf{c}|\mathbf{n}, \mathbf{V}, \alpha\right) = \mathcal{N}\left(\mathbf{c}|\alpha\mathbf{V}\mathbf{L}\mathbf{n}, \sigma^2\mathbf{I}\right). \quad (2)$$

Given an observed image value $\mathbf{c}$ and an estimate of $\mathbf{V}$ and $\mathbf{L}$ the maximum likelihood estimate of $\mathbf{n}$ is then given by

$$\mathbf{n} = \frac{\left(\mathbf{V}\mathbf{L}\right)^{-1}\mathbf{c}}{\left|\left(\mathbf{V}\mathbf{L}\right)^{-1}\mathbf{c}\right|}. \quad (3)$$

The likelihood of observing an image and normal pair $(\mathbf{c}, \mathbf{n})$ given a chromaticity defined by the matrix $\mathbf{V}$ can be found using Bayes' rule as

$$p\left(\mathbf{c}, \mathbf{n}|\mathbf{V}\right) = p\left(\mathbf{c}|\mathbf{n}, \mathbf{V}\right) p\left(\mathbf{n}|\mathbf{V}\right). \quad (4)$$

A uniform prior is assumed for the surface normals $p\left(\mathbf{n}|\mathbf{V}\right)$. We cannot express $p\left(\mathbf{c}|\mathbf{n}, \mathbf{V}\right)$ without taking the surface's albedo α into account. Since this is unknown we marginalize it out, giving

$$p\left(\mathbf{c}|\mathbf{n}, \mathbf{V}\right) = \int p\left(\mathbf{c}|\mathbf{n}, \mathbf{V}, \alpha\right) p\left(\alpha|\mathbf{n}, \mathbf{V}\right) \mathrm{d}\alpha. \quad (5)$$

We set the prior $p\left(\alpha|\mathbf{n}, \mathbf{V}\right)$ to be uniform in the range zero to one. Using (2) this gives

$$p\left(\mathbf{c}|\mathbf{n}, \mathbf{V}\right) = \int_0^1 \mathcal{N}\left(\mathbf{c}|\alpha\mathbf{V}\mathbf{L}\mathbf{n}, \sigma^2\mathbf{I}\right) \mathrm{d}\alpha. \quad (6)$$

By choosing a coordinate system such that the first axis of this new coordinate system is parallel to the line $\mathbf{V}\mathbf{L}\mathbf{n}$ this can be written as

$$p\left(\mathbf{c}|\mathbf{n}, \mathbf{V}\right) = \int_0^1 \mathcal{N}\left(\mathbf{c}_r|\begin{bmatrix} |\alpha\mathbf{V}\mathbf{L}\mathbf{n}| \\ 0 \\ 0 \end{bmatrix}, \sigma^2\mathbf{I}\right) \mathrm{d}\alpha, \quad (7)$$

where $\mathbf{c}_r = \begin{bmatrix} c_{r0} & c_{r1} & c_{r2} \end{bmatrix}^\top$ is $\mathbf{c}$ in the new rotated coordinate system. Removing all terms that do not depend on α from the integral and using $b = |\mathbf{V}\mathbf{L}\mathbf{n}|$ for compactness this can be integrated to give

$$\frac{\mathcal{N}\left(d|0, \sigma^2\right)}{2b}\left(\mathrm{Erf}\left(\frac{c_{r0}}{\sigma\sqrt{2}}\right) - \mathrm{Erf}\left(\frac{c_{r0} - b}{\sigma\sqrt{2}}\right)\right), \quad (8)$$

where

$$d^2 = c_{r0}^2 + c_{r1}^2 \tag{9}$$

and Erf() is the error function. In the original coordinate system c_{r0} and d are given by

$$c_{r0} = \frac{\mathbf{c}^\intercal \mathbf{VLn}}{|\mathbf{VLn}|} \tag{10}$$

and

$$d = |\mathbf{c} - c_{r0}\mathbf{VLn}| \,. \tag{11}$$

Intuitively c_{r0} corresponds to the distance along the line $\mathbf{VLn}$ and d to the displacement perpendicular to this line due to noise. The term containing the two error functions is approximately constant between 0 and $|\mathbf{VLn}|$ due to our uniform prior upon α and as such, for practical purposes, can be treated as a constant.

3.3. Segmentation

To perform the segmentation of a new scene into different chromaticities we construct a Markov Random Field (MRF) in which each node corresponds to a pixel in the input image and is connected to the node of each of the pixel's neighbors in a 4-neighborhood. Each node will be assigned a label $a \in 1, \ldots, N$ corresponding to one of the N chromaticities in the scene. The constraint that chromaticity should be locally constant is enforced using the Potts model for the pairwise terms in which no cost is assigned to neighboring pixels sharing a label and a cost γ is assigned for differing labels. The unary terms are given by the likelihood derived in the previous section. Given a set of N matrices, $\mathbf{V}_{a\in1,\ldots,N}$, the unary term for a pixel taking label a is given by $P(\mathbf{c}|\mathbf{n}, \mathbf{V}_a)$ where the $\mathbf{n}$ is taken from the smoothed normal map estimated from the stereo depth map and $\mathbf{c}$ is an image intensity taken from a smoothed version of the input image. Smoothing is necessary to remove high frequency variation due to fine geometric detail which the stereo algorithm cannot recover.

To ensure that the segmentation boundaries follow region boundaries closely, an edge map of the image is computed and Potts costs for edges in the graph that cross an edge in the edge map are set to $\frac{\gamma}{100}$.

Once the MRF has been built, it is solved using the tree reweighted message passing algorithm of Kolmogorov [18] and normals are estimated independently at each pixel using (3) with the relevant $\mathbf{V}_a$. This dense normal map is then combined with the low-resolution stereo depth map using the method of Nehab *et al*. [22].

4. Calibration

This section describes the calibration procedure used to estimate the parameters required for reconstruction. These can be split into two groups, the photometric parameters, N and $\mathbf{V}_{a\in1,\ldots,N}$, which need to be estimated for each scene individually and the lighting direction matrix $\mathbf{L}$, the image noise σ and the camera intrinsic and extrinsic parameters, which only need to be estimated once. In order for the photometric parameters to be estimated the scene must be held still for long enough to acquire three images under different lighting. We did not find this to be a problem in practice.

Algorithm 1 Complete calibration procedure

Require: I_r, I_g, I_b, $\mathbf{L}$, σ, stereo depth map
1: Estimate per pixel $\mathbf{V}$ as in section 4.1.
2: **for** $N = 1 : N_{max}$ **do**
3: **for** $\tau = \tau_{min} : \tau_{max}$ **do**
4: Estimate $\mathbf{V}_{a\in1,\ldots,N}$ using RANSAC (section 4.2)
5: Segment image as in section 3.3
6: Calculate BIC using (13) and this segmentation
7: **end for**
8: **end for**
9: **return** N, $\mathbf{V}_{a\in1,\ldots,N}$

Estimation of the intrinsic and extrinsic camera parameters is carried out using the standard technique of rotating and translating a checkerboard pattern in the field of view [28]. Estimation of $\mathbf{L}$ is also carried out using a standard method; rotating the same checkerboard pattern with only one light on at a time provides a set of $(\mathbf{c}, \mathbf{n})$ pairs from which $\mathbf{L}$ can be estimated using least squares, as in [12]. To estimate σ several images of a static scene under constant lighting are acquired and σ^2 is estimated as the average variance of the pixels across the images. The procedure for estimating N and $\mathbf{V}_{a\in1,\ldots,N}$ can be broken down into three parts detailed in the following sections:

1. Estimation of $\mathbf{V}$ at each pixel individually.
2. Estimation of the N dominant chromaticities, $\mathbf{V}_{a\in1,\ldots,N}$, where N is given.
3. Selection of N as a model order selection problem.

The complete procedure is outlined in Algorithm 1.

4.1. Per pixel calibration

To estimate $\mathbf{V}$ at every pixel we propose the following method. Three images are acquired, I_r, I_g and I_b, with each light being switched on in one of the images. It is assumed that scene geometry is constant across the three images. A stereo reconstruction is also performed to give a low-resolution normal map. Given this normal map and the previously computed lighting directions, each of the three images allows for an estimate of one column of the $\mathbf{V}$ matrix to be made at each pixel. For example, using I_r when only the first (red) light is on (1) reduces to

$$\mathbf{c} = \alpha \begin{bmatrix} \mathbf{v_0} & \mathbf{v_1} & \mathbf{v_2} \end{bmatrix} \begin{bmatrix} \mathbf{l_0} & 0 & 0 \end{bmatrix}^\top \mathbf{n} = \alpha \mathbf{v_0} \mathbf{l_0}^\top \mathbf{n}. \tag{12}$$

Since $\mathbf{c}$, $\mathbf{n}$ and $\mathbf{l_0}$ are known, this allows all elements of $\mathbf{v_0}$ to be calculated up to the scaling factor α, which is constant across all columns in $\mathbf{V}$. To account for the fact that the stereo normal map does not recover high frequency geometry, each of the three images are smoothed before this process is carried out.

This procedure actually recovers $\alpha\mathbf{V}$ at each pixel, not $\mathbf{V}$. As can be seen from (3) the scale of $\mathbf{V}$ is unimportant during reconstruction, so we scale each $\mathbf{V}$ matrix so that the largest $\mathbf{c}$ value it can predict given a valid normal has a value not greater than 255. We ensure that saturation does not occur in practice by adjusting the camera's exposure and gain settings.

4.2. Calibrating for multiple chromaticities

Once an individual calibration matrix has been estimated for each pixel, we wish to find both N, the number of chromaticities present in the scene, and $\mathbf{V}_{a\in1,\dots,N}$ which are the photometric properties of the N dominant chromaticities in the scene. Initially assuming that N is known, we wish to choose $\mathbf{V}_{a\in1,\dots,N}$ to explain the scene as well as possible. To do this we use a RANSAC-based approach similar to that of [11]. One of the calculated $\mathbf{V}$ matrices is chosen at random as a hypothesis and the number of pixels in the calibration scene that support it is observed. To measure support for a hypothesis an image under full multispectral lighting I_{rgb} is synthesized according to $I_{rgb} = I_r + I_g + I_b$. Using the pixel intensities $\mathbf{c}$ from this synthesized image and the previously computed normals, the likelihood of this $(\mathbf{c}, \mathbf{n})$ pair given the hypothesized $\mathbf{V}$ matrix can be calculated using (8). If the likelihood is above a threshold value τ the pixel supports the hypothesized matrix, otherwise it does not.

This is repeated a fixed number of times retaining the hypothesis with the most support each time. The final calibration matrix is then found by averaging $\mathbf{V}$ over all the pixels that supported the final hypothesis. Once the first calibration matrix has been chosen, all pixels that supported it are removed and the process is repeated to find the next most dominant chromaticity in the scene. This is repeated until N calibration matrices have been recovered.

4.3. Estimation of the number of chromaticities N

The above procedure assumes that N is already known, however this is not the case. Selection of N can be viewed as a model selection problem in which the increase in the model's ability to explain the input image by increasing N is traded off against the added complexity of the model. We wish to use an information theoretic model selection method, and to reduce the chance of overfitting we use the Bayesian Information Criterion (BIC) [23]. Once the RANSAC stage has been carried out to estimate $\mathbf{V}_{a\in1,\dots,N}$, an MRF can be solved as it would be during reconstruction so that the correct $\mathbf{V}_a$ can be used at each pixel in the image. Assuming pixel-wise independence the likelihood of the complete image is the product of the pixel likelihoods and hence the BIC score can be calculated using

$$\text{BIC} = -2\sum_{i=1}^{n} \ln P\left(\mathbf{c}_i, \mathbf{n}_i | \mathbf{V}_{a_i}\right) + mN \ln n, \qquad (13)$$

where n is the number of pixels in the image, and m is the number of additional model parameters when increasing N by one, nine in this case. The value of N that produces the lowest BIC score is chosen. In practice this process is repeated for five values of the threshold τ for each N and the lowest BIC score over all N and τ is used.

5. Experiments

5.1. Constant chromaticity calibration

While focusing on multichromatic scenes, the proposed calibration method is applicable to scenes of uniform chromaticity. To demonstrate this we reconstructed three different faces using both our method and that of [11]. Recovered results were very similar with an average difference in normal of $0.54°$. An example reconstruction is given in the top row of figure 4.

5.2. Multichromatic scenes

To demonstrate our approach on a multichromatic scene a challenging sequence involving a green and white cushion was processed. Calibration was performed and resulted in $N = 3$ being selected (two chromaticities on the cushion and another for the hands holding it). If only one chromaticity is assumed for the entire scene, the hand is recovered incorrectly, see figure 3(a), whilst using the proposed method the normals and therefore geometry are correctly estimated, see figure 3(b).

Throughout the 300 frame sequence segmentation is sufficiently accurate to give qualitatively good results. For some frames such as that in figure 3(c) segmentation fails due to strong shadowing, here failing to segment the fingers correctly. The resulting geometry produced, see figure 3(e), exhibits artifacts, but this failure is a rare occurrence.

The resulting reconstructions are best viewed dynamically in the supporting video, but several stills have been provided in figure 4 for illustration.

The stereo results shown in the second column contain little detail but give correct overall geometry. The second and third columns show results comparable to those achieved by purely photometric systems such as [11] which look convincing when viewed from close to the original viewing direction, but contain low frequency deformations that become apparent when rendered from novel viewpoints. Our combined results in the final two columns

(a) (b) (c) (d) (e)

Figure 3. **Reconstruction detail:** Reconstruction assuming (a) a single chromaticity, and (b) assuming 3 chromaticities. Note the improvement of the reconstructed hand shape. **Failure case:** (c) Input image, (d) failed segmentation, (e) resulting reconstruction.

demonstrate that this low frequency deformation has been removed while retaining high frequency detail.

The image size for these sequences is 1600×1200 and the mean running time of the complete algorithm is 16 seconds per frame, the two most time consuming parts being the stereo reconstruction in CUDA (4 seconds) and the normal field integration (9 seconds) in single-threaded C++. The mean segmentation time is approximately 2 seconds.

5.3. Quantitative analysis

In order to demonstrate the accuracy of our approach against ground truth data, a set of experiments on synthetic images was carried out. A publicly available high resolution model captured in [20] was rendered in OpenGL. The diffuse albedo recorded by [20] was applied to half of the model and the other half was rendered using three different solid colors. A uniform white specular albedo was present over the entire model. An example input image is shown in figure 5(a).

Initially no noise was added to the images and reconstruction was carried out. The errors between ground truth normals and recovered normals are shown in figure 5(c). In areas of uniform chromaticity errors are due to specular reflections or region boundaries while in the unmodified half there is a varying level of error introduced by the varying chromaticity.

Calibration was also carried out using the method of [11] with resulting normal errors shown in figure 5(b). This approach estimates the correct calibration for the unmodified portion of the model, producing similar results to the proposed method in this region, but it cannot deal with the multiple chromaticities in the scene.

If the recovered normal field is integrated then there is a large discrepancy between recovered depth and ground truth values as shown in figure 5(e) due to a slight bias in the normal estimation. Combining the depth map estimated using stereo with the normal maps greatly reduces this error as shown in figure 5(f).

To simulate image noise we added Gaussian noise with a standard deviation of 6 independently to each color channel of each pixel and repeated the above experiments. Numerical results for depth and normal errors are given in table 1.

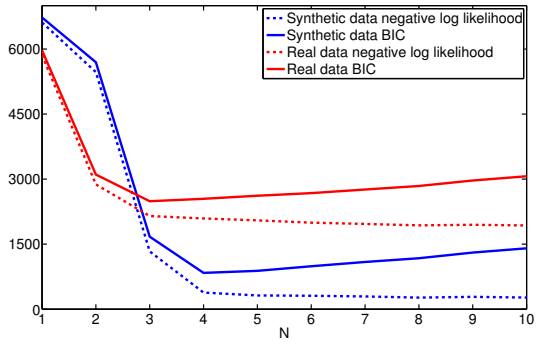


Figure 6. **Model selection**. Plots of negative log likelihood and resulting BIC value as N, the number of colors in the scene, is increased. (Blue) for synthetic face data with four major chromaticities (red) for real cushion data for which there are three major chromaticities. In both cases the plots are the average values from 100 runs and dashed lines show negative log likelihood while solid lines show the BIC value.

5.4. Estimation of the number of chromaticities N

In all of the above experiments N was estimated using model selection with the BIC criterion. Here we present results demonstrating the stability of our approach to estimating N. Figure 6 shows plots of the negative log likelihoods used in (13) and the corresponding BIC values for the synthetic face (figure 5), which has four major chromaticities, and the real cushion (figure 3) which has three major chromaticities. It can be seen that in each case the correct N is chosen. Also in both cases the rates of reduction of log likelihood decreases rapidly beyond the correct N value.

5.5. Current limitations

The proposed approach makes two major assumptions about the scene, firstly that it contains a small number of distinct chromaticities and secondly that it is well approximated by a Lambertian reflectance model. Reconstructions of scenes which break these assumptions will contain artifacts. Also reconstruction of dark surfaces is noisy due to the low amount of reflected light.

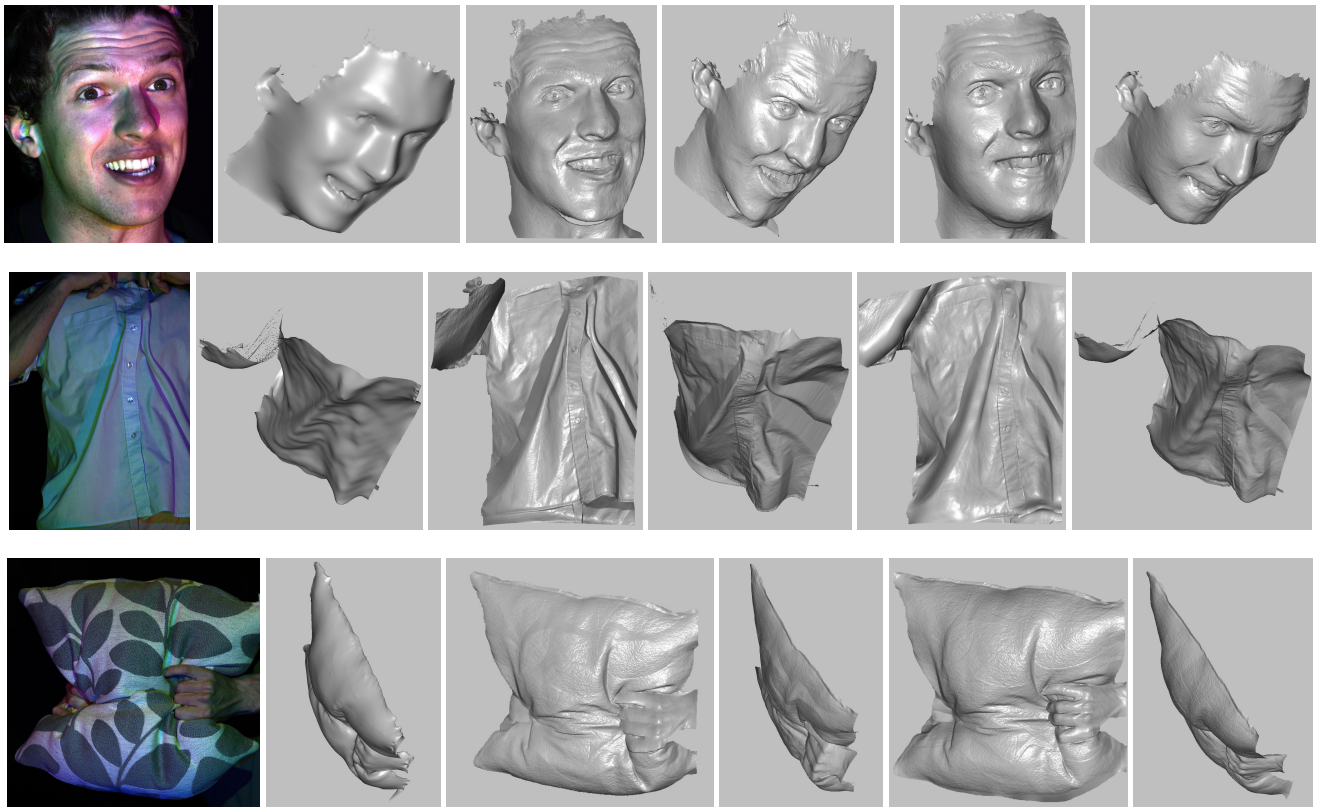


Figure 4. **Reconstructions of real sequences**. From left to right in each row: Input image, stereo reconstruction, integrated normal field, novel view of integrated normal field, final result once stereo information included, same novel view of final result. While integrating the normal field gives reasonable results when viewed frontally, low frequency deformations are visible when viewed from novel angles. These deformations are removed in the final result using low frequency information from stereo. The top row uses $N = 1$, the middle row $N = 2$ and the final row $N = 3$ chromaticities in the scene, which are found automatically using model selection.

		Stereo only	Normals only Calibration by [11]	Normals only New calibration	Stereo + normals New calibration
No noise	Normal error (°). Mean (std dev)	11.8 (10.3)	22.6 (23.2)	3.97 (5.23)	**3.26 (4.07)**
	Depth error (mm). Mean (std dev)	0.39 (1.18)	10.3 (10.9)	6.83 (5.49)	**0.37 (1.23)**
Noise $\sigma = 6$	Normal error (°). Mean (std dev)	11.9 (10.4)	25.2 (24.1)	9.06 (6.06)	**8.37 (5.62)**
	Depth error (mm). Mean (std dev)	0.40 (1.21)	10.3 (10.9)	6.86 (5.51)	**0.38 (1.27)**

Table 1. **Errors on ground truth data**. Assuming constant chromaticity as in [11] leads to large errors. Whilst the stereo data provides accurate depths, the geometry is over-smoothed, making normal estimation inaccurate. The proposed reconstruction method accurately estimates normal directions, but the addition of the stereo data is still necessary to remove low frequency bias in depth results.

6. Conclusions

We have presented a system for applying multispectral photometric stereo to scenes containing multiple chromaticities by making use of multiview stereo reconstruction. A novel calibration technique was demonstrated that allows photometric properties to be estimated at each pixel in a calibration scene. It was shown that automatic estimation of the number of chromaticities in such a scene can be carried out using a model selection approach. Given such a calibration we are able to segment new images into regions of constant chromaticity and produce dense normal maps.

References

[1] N. Ahmed, C. Theobalt, P. Dobrev, H. Seidel, and S. Thrun. Robust fusion of dynamic shape and normal capture for high-quality reconstruction of time-varying geometry. In *CVPR*, June 2008. 2

[2] O. Alexander, M. Rogers, W. Lambeth, M. Chiang, and P. Debevec. Creating a photoreal digital actor: The Digital Emily project. *Conference for Visual Media Production*, pages 69–80, 2009. 2

[3] R. Anderson, B. Stenger, and R. Cipolla. Augmenting depth camera output using photometric stereo. In *MVA*, 2011. 1

[4] T. Beeler, B. Bickel, P. Beardsley, and B. Sumner. High-

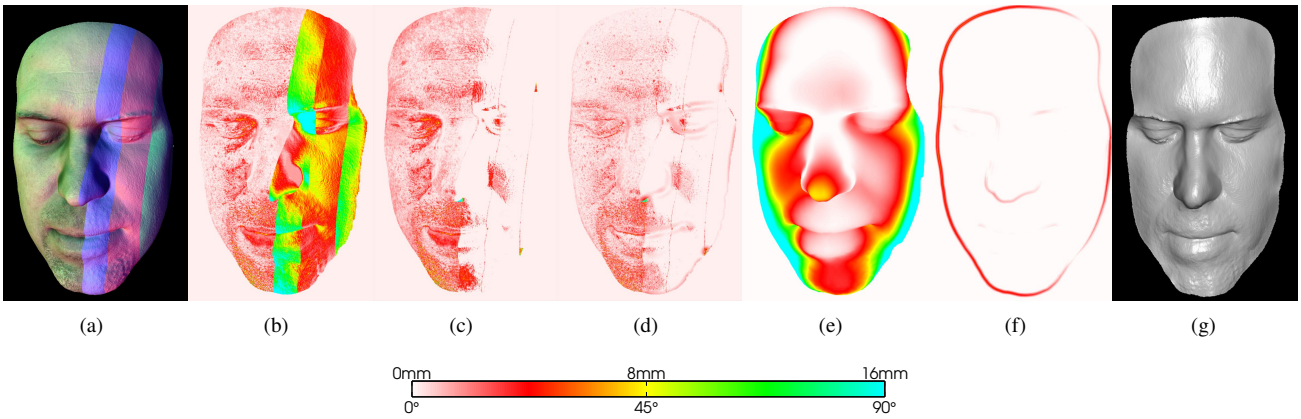


Figure 5. **Quantitative analysis on rendered data**. (a) Input image generated by coloring and rendering the mesh from [20], (b) normal errors when calibrating using [11], (c) normal errors using proposed method, (d) normal errors after addition of low frequency data, (e) depth errors after integrating normal field, (f) depth errors after adding low frequency data, (g) final reconstruction.

quality single-shot capture of facial geometry. *ACM SIGGRAPH*, 2010. 2

[5] D. Bradley, W. Heidrich, T. Popa, and A. Sheffer. High resolution passive facial performance capture. *SIGGRAPH*, 2010. 2

[6] G. Brostow, C. Hernández, G. Vogiatzis, B. Stenger, and R. Cipolla. Video normals from colored lights. *PAMI*, 2011. 1

[7] B. De Decker, J. Kautz, T. Mertens, and P. Bekaert. Capturing multiple illumination conditions using time and color multiplexing. *CVPR*, pages 2536–2543, June 2009. 1

[8] M. Drew and L. Kontsevich. Closed-form attitude determination under spectrally varying illumination. In *CVPR*, pages 985 –990, June 1994. 1, 2, 3

[9] Y. Furukawa and J. Ponce. Dense 3D motion capture for human faces. In *CVPR*, 2009. 2

[10] Y. Furukawa and J. Ponce. Accurate, dense, and robust multiview stereopsis. *PAMI*, 32(8):1362–76, Aug. 2010. 2

[11] C. Hernández and G. Vogiatzis. Self-calibrating a real-time monocular 3d facial capture system. *3DPVT*, 2010. 1, 5, 6, 7, 8

[12] C. Hernández, G. Vogiatzis, G. J. Brostow, B. Stenger, and R. Cipolla. Non-rigid Photometric Stereo with Colored Lights. *ICCV*, pages 1–8, October 2007. 4

[13] K. Ikeuchi. Determining a depth map using a dual photometric stereo. *International Journal of Robotics Research*, 6(1):1–11, 1987. 2

[14] Z. Janko, A. Delaunoy, and E. Prados. Colour dynamic photometric stereo for textured surfaces. In *ACCV*, 2010. 1

[15] M. K. Johnson and E. H. Adelson. Retrographic sensing for the measurement of surface texture and shape. In *CVPR*, pages 1070–1077, 2009. 1

[16] H. Kim, B. Wilburn, and M. Ben-Ezra. Photometric Stereo for Dynamic Surface Orientations. *ECCV*, pages 59–72, 2010. 1

[17] M. Klaudiny, A. Hilton, and J. Edge. High-detail 3D capture of facial performance. *3DPVT*, 2010. 1, 2

[18] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *PAMI*, 28(10), 2006. 4

[19] L. Kontsevich, A. Petrov, and I. Vergelskaya. Reconstruction of shape from shading in color images. *Journal of the Optical Society of America A*, 11(3):1047, Mar. 1994. 1

[20] W. Ma, T. Hawkins, P. Peers, C. Chabert, M. Weiss, and P. Debevec. Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. *In Eurographics Symposium on Rendering*, pages 183–194, September 2007. 6, 8

[21] W. Ma, A. Jones, J. Chiang, T. Hawkins, S. Frederiksen, P. Peers, M. Vukovic, M. Ouhyoung, and P. Debevec. Facial performance synthesis using deformation-driven polynomial displacement maps. *ACM SIGGRAPH*, 27(5), 2008. 2

[22] D. Nehab, S. Rusinkiewicz, J. Davis, and R. Ramamoorthi. Efficiently combining positions and normals for precise 3D geometry. *ACM SIGGRAPH*, 24(3):536–543, 2005. 2, 4

[23] P. Stoica and Y. Selen. Model-order selection: a review of information criterion rules. *Signal Processing Magazine, IEEE*, 21(4):36 – 47, 2004. 5

[24] D. Vlasic, P. Peers, I. Baran, P. Debevec, J. Popović, S. Rusinkiewicz, and W. Matusik. Dynamic shape capture using multi-view photometric stereo. In *SIGGRAPH Asia*, pages 1–11, 2009. 2

[25] T. Weise, H. Li, L. Van Gool, and M. Pauly. Face/Off: live facial puppetry. In *ACM SIGGRAPH*, pages 7–16, 2009. 2

[26] R. J. Woodham. Gradient and Curvature from Photometric Stereo Including Local Confidence Estimation. *Journal of the Optical Society of America*, (11):3050–3068, 1994. 1

[27] L. Zhang, N. Snavely, B. Curless, and S. Seitz. Spacetime faces: high resolution capture for modeling and animation. *ACM SIGGRAPH*, 23(3):546–556, 2004. 2

[28] Z. Zhang. A flexible new technique for camera calibration. *PAMI*, 22(11):1330–1334, 2000. 4