

Reflectance Transfer for Material Editing and Relighting

Peter Stroia-Williams*, Adrian Hilton*, and Oliver Grau †

Centre for Vision, Speech and Signal Processing
The University of Surrey
Guildford, GU2 7XH

phone, email: peter.williams@surrey.ac.uk

www: www.ee.surrey.ac.uk/CVSSP/home

BBC Research and Development

Kingswood Warren
Tadworth, KT20 6NP

phone, email: oliver.grau@bbc.co.uk

www: .bbc.co.uk/rd

Abstract

We present a new approach to diffuse reflectance estimation for dynamic scenes.

Non-parametric image statistics are used to transfer reflectance properties from a static example set to a dynamic image sequence.

The approach allows diffuse reflectance estimation for surface materials with inhomogeneous appearance, such as those which commonly occur with patterned or textured clothing. Material editing is also possible by transferring edited reflectance properties.

Material reflectance properties are initially estimated from static images of the subject under multiple directional illuminations using photometric stereo. The estimated reflectance together with the corresponding image under uniform ambient illumination form a prior set of reference material observations.

Material reflectance properties are then estimated for video sequences of a moving person captured under uniform ambient illumination by matching the observed local image statistics to the reference observations. Results demonstrate that the transfer of reflectance properties enables estimation of the dynamic surface normals and subsequent relighting combined with material editing. This approach overcomes limitations of previous work on material transfer and relighting of dynamic scenes which was limited to surfaces with regions of homogeneous reflectance. We evaluate our approach for relighting 3D model sequences reconstructed from multiple view video. Comparison to previous model relighting demonstrates improved reproduction of detailed texture and shape dynamics.

Keywords: Relighting, Reflectance Transfer, Material Editing, Shape From Shading, Non-parametric Statistics, GPGPU

Digital Peer Publishing Licence

Any party may pass on this Work by electronic means and make it available for download under the terms and conditions of the current version of the Digital Peer Publishing Licence (DPPL). The text of the licence may be accessed and retrieved via Internet at <http://www.dipp.nrw.de/>.

First presented at the 5th European Conference on Visual Media Production 2008, extended and revised for JVRB

1 Introduction

In this work the objective is to relight surfaces with unknown non-uniform albedo and unrestricted scene illumination. To achieve this objective surface reflectance properties are estimated from initial observations of the subject under multiple illumination conditions using photometric stereo. The estimated reflectance properties are then transferred to image se-

quence observations of a dynamic scene by matching the observations in the static scene. The same transfer mechanism is also used to transfer edited reflectance properties allowing both material editing and relighting.

Previous work on relighting can be grouped by their capture requirements. The most established group requires a static subject. This allows the subject to be captured under multiple illumination conditions, giving enough observations for reflectance and surface orientation to be fully constrained by the data. Another group of work uses high speed cameras and lighting to multiplex lighting conditions across video giving an approximation of the static case. The final group is less restrictive in its capture requirements being able to capture relightable video with a single illumination condition. The work presented here is part of the final group. Previous work on relighting and its relation to this work is discussed in detail in section 2. This work exploits a static method in the capture of reflectance observations for reference sets. Following ideas from recent work, non-parametric statistics are used to transfer reflectance observations from these reference sets across video. Related work using non-parametric statistics is also discussed in section 2.

Figure 1 gives an overview of the whole system. The specific static reflectance capture method we apply is described in Section 3. As shown in Figure 1 the reflectance observations captured are joined by an image captured under the same illumination conditions as the image sequence to form a reference set. For material editing the reflectance image is manually changed to create an edited reflectance image, which also forms a part of the reference set. The approach for reflectance transfer using the reference set and a target video is presented in Section 4. The final methodology sections 5 and 6 describe the material editing and relighting approaches based on the reflectance transfer results. Results and a comparative evaluation with existing approaches for relighting dynamic scenes of people is presented in Section 7 followed by conclusions in Section 8.

2 Related Work

Over the past decade there has been extensive interest in performance reconstruction from multiple view video for *free-viewpoint* 3D video replay. Research has primarily focused on accurate reconstruction of shape and high-quality rendering of novel views of the

captured performance. Reuse for entertainment production in film, broadcast or games often requires relighting to composite the captured performance within a novel scene. Photo-realistic relighting of dynamic scenes remains a challenging open-problem. In this section we review research related to the problem in the estimation of surface reflectance properties and appearance transfer to manipulate dynamic scene appearance.

2.1 Reflectance Estimation and Relighting Dynamic Scenes

Estimation of surface reflectance properties from image observations is a classical problem in computer vision. In general the observed surface appearance depends on the illumination, surface reflectance, surface shape and local orientation (normal). Reconstruction of the surface reflectance is an ill-conditioned inverse problem which requires accurate knowledge of both surface orientation and illumination. Photometric stereo [BP03] allows the estimation of the surface albedo or diffuse component for Lambertian surfaces from three or more images of the object taken under different point light sources with known position and intensity. Acquisition of the surface with multiple light sources requires a static scene to recover the albedo independently for each pixel. Photometric stereo has recently been used to recover per pixel surface normals using simultaneous acquisition with multiple coloured lights from different directions [VBS⁺07]. This approach is limited to Lambertian surfaces with uniform albedo restricting its application to real scenes. Estimation of both diffuse and specular surface reflectance from multiple view capture has been achieved using approximate temporal correspondence to obtain multiple observations of surface regions with uniform albedo over time [TAL⁺07]. However, surface tracking is not sufficiently effective for this approach without using simplified models that lead to significant biasing in the large-scale shape. Also this approach requires clustering of surface observations limiting the approach to materials with homogeneous colour regions.

Recovery of parametric models of surface reflectance properties for static objects has been achieved using an accurate prior measurement of surface shape obtained from range scans [SWI97]. For dynamic scenes accurate reconstruction of fine surface detail from multiple view images or video-rate range

scans remains an open problem. Fine detail such as creases in clothing is typically not reproduced in surface reconstructions of a person [SH07] resulting in surface normals with insufficient accuracy to recover the reflectance. Relighting of large scale structures assuming Lambertian reflectance can be achieved by estimating the albedo given inaccurate measurements of shape and orientation [Gra06].

Previous approaches for capturing reflectance of dynamic deformable subjects have made use of illumination maps and surface shape data allowing the surface reflectance to be calculated directly from the diffuse reflectance model [CH06, Gra06]. Such an approach depends on the accuracy of the illumination map and surface shape data. In particular estimation of accurate surface normals is critical to recovery of surface reflectance properties. The resultant reflectance map may be contaminated by any shading not accounted for by the illumination model and surface shape data. In Section 7.2 we compare our approach with de-lighting as part of a relightable model capture framework.

These approaches enable large scale relighting of dynamic scenes but do not relight fine detail such as wrinkles in clothing. An alternative approach to relighting dynamic scenes recently proposed uses detailed normal estimates from surface shading [CVH07]. This approach assumes surface regions with uniform Lambertian reflectance which are segmented to obtain a single albedo estimate for each region. Given the albedo estimate the shading within each region is used to estimate the local surface orientation and illumination direction. Estimation is regularised using a prior coarse model of surface shape reconstructed from multiple views. The approach achieves realistic relighting of clothing detail such as creases but is limited to clothing with uniform appearance.

Detailed dynamic shape has been successfully captured using standard studio lighting and cameras using shape-from-shading based approaches to estimate the surface normal [CVH07, HZ07]. However these approaches are limited to cloth materials with uniform colour regions. The reflectance estimation approach introduced in this paper allows estimating of surface shading for materials with non-uniform reflectance properties. Subsequent estimation of surface normals and transfer of material reflectance properties allows relighting of dynamic scenes with textured and patterned surface appearance.

Image-based relighting uses images of the scene captured under a large number of illumination condi-

tions to produce images with complex illuminations by combining the captured images. This approach has been applied to dynamic scenes using a light-stage with high-speed (4000Hz) switching of illumination sources and image capture [CEJ⁺06]. Results demonstrate photo-realistic video sequences of performers under novel illumination conditions. The use of specialist lighting and high-frame rate cameras limits the wide-spread application of this approach.

In [PTMD07] Pieter Peers et al transfer quotient images representing the change in appearance from a performance illumination to a target illumination. Using a user guided optical flow approach the quotient image is transferred across video to match pose changes or even different actors. At the large scale level the approach appears accurate but for changes in fine surface detail inaccuracies become visible. The approach presented here does not require any user input in its reflectance transfer stage. Our reflectance capture method makes use of image based non-parametric statistics. Non-parametric statistics have recently been used to represent the local image appearance structure in a number of contexts: image analogies [HJO⁺01]; image infilling [EL99]; and view-interpolation [WF05]. Nonparametric statistics provide a general representation of the local structure observed in an image region without a prior model or assumptions on appearance, allowing indexing to identify similar regions. In this work non-parametric statistics are used to represent the reflectance properties of a static scene captured under multiple illumination conditions. The reflectance properties are then transferred to novel image observations in video sequences by searching for reference image regions with similar local appearance statistics. This potentially allows us to bring the accuracy of the static only methods to dynamic sequences. Also the same transfer process can be used to transfer edited reflectance values across the sequence allowing material editing.

In [CVH07] detailed surface normals are estimated by refining the results of a multi-view 3D reconstruction under the assumption of uniform diffuse regions being lit under a single point light source. The uniform colour regions are segmented from the video using a mean-shift based spatio-temporal approach, giving a temporal consistent segmentation. For each of the regions a separate process of light source position estimation and normal estimation is performed using a shape from shading based approach. We extend this approach by replacing the segmentation with our re-

flectance estimation approach, allowing it to be applied to subjects with complex texture patterns.

2.2 Appearance Transfer

Manipulation of dynamic scene appearance in images and video has been investigated for editing texture appearance [FH04, FH06] and material properties [KRFB06]. These approaches use normal from shading to estimate the local surface orientation to modify the image appearance. Fang and Hart [FH06] transfer novel texture map appearance into video sequences using temporal tracking and shape-from-shading to estimate orientation. However, these approaches do not address the problem of relighting in images.

In this paper we propose a technique to transfer the material properties from reference images with known albedo to a video sequence. Lambertian reflectance properties for reference images are estimated using photometric stereo on a set of images captured in a static pose. Mapping of reflectance properties from the reference image to the novel image sequence of a dynamic scene is based on matching local image structure. Block based representation and matching of image structure has previously been used for image synthesis, infilling and view-interpolation [HJO⁺01, EF01, FWZ03].

Local image blocks provide a non-parametric representation of the image statistics which can be used as a prior to constrain the synthesis of novel image patches. In this paper we demonstrate that an analogous approach can be used to transfer material properties from a reference image to a video sequence captured under the same illumination conditions.

3 Initial Estimation of Material Properties

Photometric stereo [BP03] is used to estimate the surface normal and albedo of the scene assuming Lambertian reflectance. For each reference pose p of the subject we capture a set of n images $\{\mathbf{I}_i : 1 \leq i \leq n\}$, where $n \geq 3$. Each under a different known point light source illumination with directions $l_i = (l_i^x, l_i^y, l_i^z)$ together with an image \mathbf{R}_p under the same illumination conditions as the dynamic capture. Photometric stereo gives an estimate albedo image \mathbf{A}_p and normal image \mathbf{N}_p . Applying photometric stereo for each reference pose gives a set $I = \{\mathbf{R}_p, \mathbf{A}_p, \mathbf{N}_p : 1 \leq p \leq n_p\}$ of reference images. Where n_p is the number of

poses. This forms the basis for transferring material properties to the dynamic scene observed under the same illumination conditions as the reference image set \mathbf{R}_p . Our photometric stereo method follows Barskys colour photometric stereo method [BP03]. The colour radiance value $r_i = (L_i^r, L_i^g, L_i^b)$ observed under illumination directions $l_i = (l_i^x, l_i^y, l_i^z)$ must be split into colour $c = (c^r, c^g, c^b)$ and intensity components z_i such that the error between the left hand side and right hand side of Equation 1 is minimised. We do this using a least squares approach.

$$\mathbb{I} = \begin{bmatrix} L_1^r & L_1^g & L_1^b \\ \vdots & \vdots & \vdots \\ L_n^r & L_n^g & L_n^b \end{bmatrix} \simeq \begin{bmatrix} z_1 c^r & z_1 c^g & z_1 c^b \\ \vdots & \vdots & \vdots \\ Z_n c^r & Z_n c^g & z_n c^b \end{bmatrix} \quad (1)$$

Then using

$$[L] = \begin{bmatrix} l_1^x & l_1^y & l_1^z \\ \vdots & \vdots & \vdots \\ l_n^x & l_n^y & l_n^z \end{bmatrix} \quad (2)$$

and

$$\mathbf{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} \quad (3)$$

greyscale photometric stereo can be applied to the intensity part.

$$\rho \mathbf{n} = \left([L]^T [L] \right)^{-1} [L]^T \mathbf{z} \quad (4)$$

Where n is the surface normal, $\rho = \|\rho\|$ is the greyscale albedo and the colour albedo ρ is found as bellow.

$$\rho = \rho c \quad (5)$$

In order to capture this reflectance set we used 5 standard studio spot lights and a studio quality HD video camera along with the lighting setup for the dynamic capture. The spot lights were arranged with 4 roughly in a circle around a video camera and the 5th alongside the camera. The directions of the lights to the centre of the capture volume relative to the camera were calibrated using an image of the setup captured with a Spheron HDR camera from the centre of the capture volume. Any camera with known intrinsics would be suitable. The centres of the spot lights

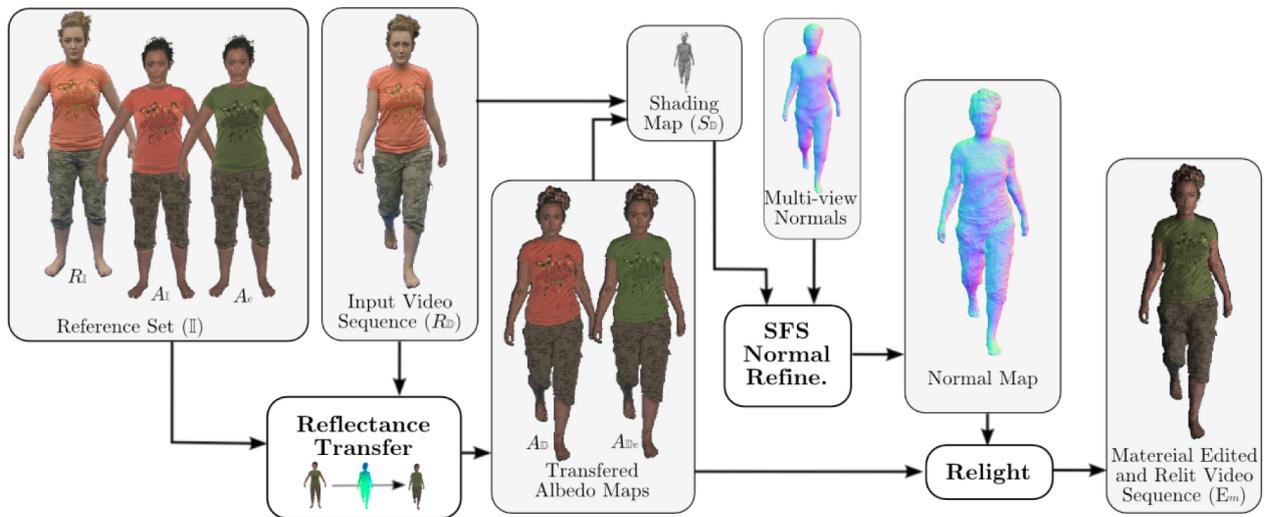


Figure 1: Process Overview

were manually selected from this image allowing the directions from the centre of the capture volume to be calculated. The intensity of the spot lights were calibrated using a diffuse sphere placed in the centre of the capture volume. The maximum intensity for each spot light was observed through the video camera and the intensities were adjusted to match that of the weakest. Previously we have tried using the spot lights at their maximum settings and compensating in the photometric stereo calculation. However such an approach is less accurate due to the less effective use of the cameras dynamic range.

To capture the required images the actor was asked to stand still in the centre of the capture volume as the 5 spot lights and dynamic capture studio lights cycled through, giving 5 illumination conditions for the photometric stereo calculation and the illumination condition under which the dynamic sequences were to be captured. The total sequence takes less than 10 seconds to capture with the limiting factor being the fade in time for the spot lights.

4 Reflectance Transfer

In this section a method for transferring reflectance from a static reference set to a dynamic sequence is described. Section 4.1 describes the overall algorithm. Then in Section 4.2 the details of the matching function used by the algorithm are given. Section 4.3 describes possible computation approaches comparing

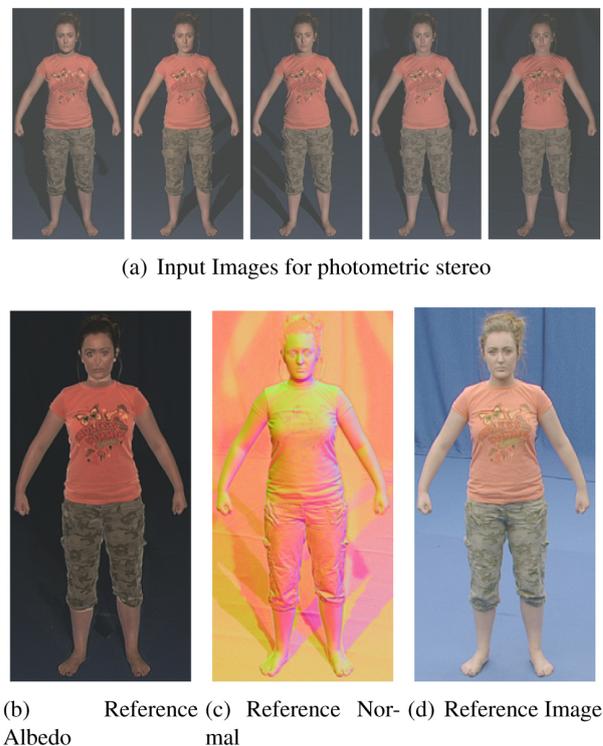


Figure 2: Example reference set.

an optimized CPU based approach with a GPU implementation.

4.1 Example-based Albedo Transfer

Estimation of the surface albedo for dynamic scenes with unknown surface orientation and illumination is based on the transfer of albedo estimates from the reference examples of the same subjects which include images captured under the same illumination conditions as the dynamic scene. Transfer is based on the similarity of local image structure in the dynamic and reference images. This approach is similar to the non-parametric representation of image structure used in previous work for image analogies, infilling and view interpolation [HJO⁺01, EL99, WF05].

In this work for each pixel position (i, j) and time t in the dynamic sequence, $\mathbf{R}_{\mathbb{D}}(i, j, t)$ the best match index is estimated by finding the region in the reference image set $\mathbf{R}_p(u, v)$ that minimises the difference in local image appearance, $d()$ as bellow. This best match index is stored giving a transfer map $\mathbf{T}(i, j, t)$ for the sequence. This is latter used to transfer albedo values from the reference set across the sequence.

$$\mathbf{T}(i, j, t) = (u, v) : \arg \min_{u, v, p} d(\mathbf{R}_{\mathbb{D}}(i, j, t), \mathbf{R}_p(u, v)) : (\mathbf{R}_p) \in \mathbb{I}, p \in [1, N_I], (u, v) \in [1, N]$$

Where $d(\mathbf{I}_1(i, j), \mathbf{I}_2(u, v))$ is a measure of the local image difference for a region around the pixel (i, j) in image \mathbf{I}_1 and pixel (u, v) in image \mathbf{I}_2 . The underlying assumption of this approach is that the reference image set contains examples of the reference material with similar local illumination, orientation and shape. A single reference image of clothing acquired under the same illumination as the dynamic scene commonly contains multiple examples of the same material with different orientations and local surface deformation providing a set of exemplars for matching. A Lambertian reflectance model is assumed. In the case of non-Lambertian materials the specular component will be treated as part of the diffuse appearance requiring an increased number of exemplars for matching. The algorithm for albedo transfer is presented in Algorithm 1. For each pixel in the observed dynamic scene $\mathbf{R}_{\mathbb{D}}(i, j, t)$ we search for the exemplar with minimum difference in appearance from the reference image set as defined by equation 6. The position of this exemplar is then stored in the transfer map $\mathbf{T}(i, j, t)$, giving a

look-up for every pixel in the sequence into the reference set.

Input : Reference Studio lit image $-\mathbf{R}_p$,
Observed Video Frame $-\mathbf{I}_{\mathbb{D}}$
Output: Transfer map for video frame $-\mathbf{T}$
for $i \leftarrow$ each pixel index in $\mathbf{I}_{\mathbb{D}}$ **do**
 $bestMatchIndex \leftarrow$
 $bestMatchSearch(i, \mathbf{I}_{\mathbb{D}}, \mathbf{R}_p)$
 $\mathbf{T}_{[i]} \leftarrow bestMatchIndex$
end

Algorithm 1: Example based transfer map calculation

4.2 Non-parametric matching

The function `bestMatchSearch` in Algorithm 1 which evaluates the exemplar with minimum difference $d()$ from all exemplars in the reference set \mathbb{I} is critical to the quality of the results. As the actor moves around the scene the surface illumination will change due to variations in illumination in the scene, changes in surface orientation within the illumination field, self shadowing and inter-reflection. We seek to compensate for these effects as well as those due to noise through our choice of best match metric.

We use a normalised intensity sum of squared distance (NYSSD) metric which scales the pixel radiance colour values by the average intensity within an $n \times n$ window around the pixel (i, j) . This metric removes global intensity difference between the observed and reference image windows to measure the difference in the local colour image structure. Normalising by the average intensity removes the effect of large scale illumination variations across the entire window. This gives a degree of illumination invariance but at the cost of reduced sensitivity to albedo variations as illumination and albedo intensity variation are inseparable. Under the assumption of a smooth surface and constant illumination this gives good illumination invariance for textured surfaces as the texture gives support to the match hypothesis. We find this to be a good compromise as albedo is more likely to have local variations than the illumination field.

The NYSSD metric for an $n \times n$ window is applied

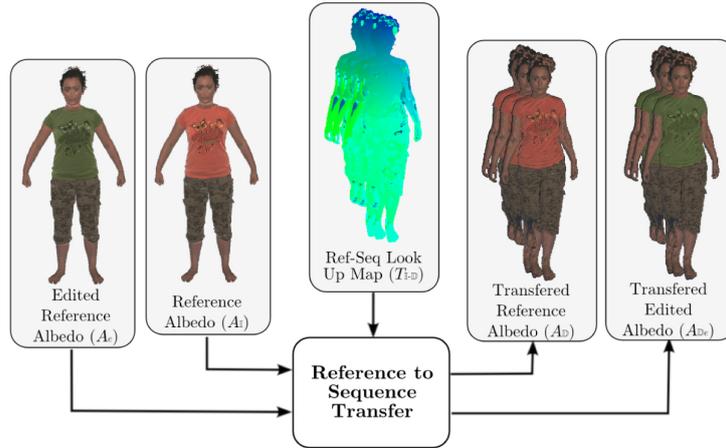


Figure 3: Reference to Sequence Transfer

in RGB colour space as follows:

$$d(\mathbf{P}_1, \mathbf{P}_2) = \sum_{x=-\frac{n}{2}}^{\frac{n}{2}} \sum_{y=-\frac{n}{2}}^{\frac{n}{2}} \left\| \frac{\mathbf{P}_1(x, y)}{\mu(\mathbf{P}_1)} - \frac{\mathbf{P}_2(x, y)}{\mu(\mathbf{P}_2)} \right\|^2 \quad (7)$$

Where \mathbf{P}_i is an $n \times n$ image patch, $\mathbf{P}_i(x, y)$ is the RGB colour vector at the position (x, y) in the patch, and $\mu(\mathbf{P}_i)$ is the mean intensity of the patch.

In order to better locate the patches we apply a Gaussian weighting to emphasise the centre of the patch in the matching function.

4.3 Nearest Neighbour Search

The search for the exemplar patch which minimises the NYSSD distance $d_{NYSSD}()$, defined by equation 7, for each pixel in the observed image has a computational complexity of $O(n^2 NN_I)$ where n is the patch width, N is number of pixels in the reference images and N_I is the number of reference images. Full search on a conventional CPU is therefore prohibitively expensive. The following explores two options to overcome this, the optimised ANN (Approximate Nearest Neighbour) library [AMN⁺98] and a GPU full search implementation. the reference image into a search structure.

For our case ANN finds the nearest neighbour with an error $\leq e$ with a computational complexity of $O(c_e \log(n^2 NN_I))$, where c_e is a constant depending on e . In order to achieve this it must first pre-process the reference image into a search structure. This requires $O(n^2 NN_I \log NN_I)$ time and $O(n^2 NN_I)$

space. The ANN library is limited to L^1 , L^2 and L^∞ norms. In order to match with NYSSD using ANN the L^2 norm is applied to intensity normalised patches.

Similar to [GDB08] our GPU implementation is written in CUDA. Each patch match function calculation is assigned to an individual thread, giving one thread per reference image pixel, see Algorithm 2. Varying from [GDB08] this work only needs the best match so it is possible to use an optimised function from the cublas library shipped with CUDA to find the best score from the raw patch comparison results.

Input : Reference Image R , Key Patch K_p

Output: Match score image M

for $i \leftarrow$ each pixel index in R **do**

InParallel

$M_{[i]} \leftarrow \text{SSD}(i, R, K_p)$

end

Algorithm 2: CUDA SSD computation

GPU's provide very high memory bandwidth and floating point computation intensity compared to standard CPU's. To make use of this there is additional work in managing memory and fitting within the per thread resource restrictions. Our implementation is currently limited by register usage, with 20 registers needed per thread limiting the number of threads that can run simultaneously to half of what it could be otherwise. Most of the registers are used as part of the memory management, so a substantial speed-up could potentially be gained with a more register efficient approach.

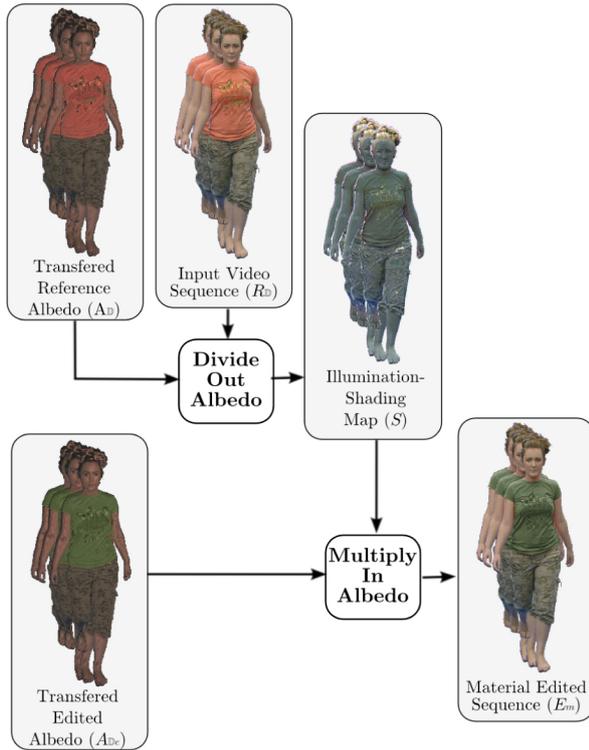


Figure 4: Material Editing Overview

Comparison

In order to compare the methods a Nvidia GeForce 8800 GTS was used for the GPU method and the CPU approaches were run on a 3.0 GHz Intel Xeon processor.

CUDA and a simple CPU full search implementation require $O(N_I)$ space to store the match function score, while ANN requires $O(nNN_I)$ space. In practice for a 311 by 679 image and 15 by 15 patches we find ANN requires in the order of GBs of memory as opposed to hundreds of MBs needed by a simple CPU or the CUDA implementation.

Figure 5 gives a comparison of computation time between a simple full search CPU implementation, ANN and our GPU full search implementation. The preprocessing stage for ANN is not included in the times whilst data transfer to the GPU is. Table 1 shows the speed-up of our GPU implementation relative to ANN. For smaller patch widths the transfer of data to the graphic card memory adds a significant delay to the calculation with a single pixel patch being slower than with ANN. From patch widths of 7 and above the speed-up starts to decline from its peak of over 20 times as the lower computational complexity of ANN

catches up with the wide parallelisation of the GPU.

5 Material Editing

Having calculated the transfer map from the reference pose to the dynamic sequence static albedo maps can be transferred across the sequence as illustrated in Figure 3. Considering each frame in a pixelwise fashion the sequence is divided by the transferred albedo leaving the illumination-shading s component, as shown in Equation 9.

$$s = \frac{r}{\rho} \quad (8)$$

Where, throughout this work we define ab to mean component-wise multiplication of the vectors and a/b to mean component-wise division of the vectors.

$$ab = \begin{bmatrix} a_0 \times b_0 \\ \vdots \\ a_i \times b_i \end{bmatrix}, \quad \frac{a}{b} = \begin{bmatrix} a_0/b_0 \\ \vdots \\ a_i/b_i \end{bmatrix} \quad (9)$$

The edited albedo value is then multiplied by the illumination-shading map giving the material edited result r_e with the original shading preserved.

$$r_e = s\rho_e \quad (10)$$

Applying this to each pixel in the full sequence gives the material edited image R_e . Figure 4 illustrates the process.

6 Relighting

An overview of the relighting process can be found in Figure 6. Having obtained an albedo map for a frame of the dynamic sequence a normal map is required for relighting. We follow the approach of Csakany [CVH07] for normal estimation. Instead of following the assumption of uniform colour regions we are able to apply the method on the illumination shading described in the previous section. Here we use the notation $S(i, j, t)$ to mean the illumination shading component value at position (i, j) and time t in the sequence.

Equation 12 forms the basis of Csakany's approach [CVH07] It is based on the assumption that the surface has uniform Lambertian reflectance, is lit by a known point light source from the direction \mathbf{L} and has a continuous convex form.

$$G(i, j, t) = \nabla S(i, j, t) - (\nabla S(i, j, t) \cdot \mathbf{L})\mathbf{L} \quad (11)$$

Where $\nabla S(i, j, t)$ is the gradient of S at (i, j, t) .

$$\mathbf{n}(i, j, t) = \frac{S(i, j, t)}{S_{\max}} \mathbf{L} + \sqrt{1 - \left(\frac{S(i, j, t)}{S_{\max}}\right)^2} \frac{G(i, j, t)}{\|G(i, j, t)\|} \quad (12)$$

The algorithm is a 3 stage process. The first stage calculates the effective light source position by estimating the normals with the above shape from shading formula and finding the rotation between these and the normals of a prior model. Next, the normals are re-estimated using the new light position before being merged with the prior model using a weighted average.

We have briefly looked into transferring the surface normals in the same way as our albedo estimation. Initial tests produced noisy results due to a combination of lack of spatial coherence in the transfer map $T_{\mathbb{I} \rightarrow \mathbb{D}}$ and the more complex relationship between surface shape and appearance not being effectively modelled by the reference set \mathbb{I} .

For relighting point light sources and Lambertian reflectance are assumed allowing the relit image to be calculated directly from the normal map, albedo map and light position as follows:

$$E_r(i, j, t) = k_a A(i, j, t) + \sum_{m=1}^n A(i, j, t,) (n_{ijt} \cdot \mathbf{L}_m) \quad (13)$$

Where $E_r(i, j, t)$ is the relit image, k_a is the amount of ambient light, L_m is the m th light source direction and n is the number of light sources. The albedo image used can either be the transferred original albedo or the transferred edited albedo, allowing combined material editing and relighting.

7 Results and Evaluation

The method presented here was integrated into a capture session using the multi-view 3D capture method presented in [SH07]. The multi-view 3D capture used 8 high-definition studio cameras and 9 standard-definition cameras running at 25fps; evenly spaced at a height of 2.5m around the capture volume. We set up our photometric stereo setup as described in Section 3 around one of the high-definition cameras. Several performances by an actor were captured with 2 different costumes. Reference sets for the two costumes

were captured consisting of a single pose each, though the use of more poses could be expected to give improved results.

All the results presented here for our method were generated using a patch width of 15 pixels for the best match search; corresponding to a square approximately $2cm^2$ across on a surface facing the camera. This was found to be a reasonable compromise between result quality and computation time. Section 7.2 presents a comparative evaluation of our approach with two other methods, before which we show some sample results.

7.1 Example Results

Figure 7 shows results using varying patch match functions, relit with a point light source along with the original images from the source video. The sequence was generated using the reference set shown in Figure 2. The sequence shows the NYSSD method to be the most stable across the sequence due to its robustness against illumination variations. Figure 8 shows that the transfer is spatially consistent in textured areas and inconsistent in uniform regions. This results in a valid albedo transfer as the uniform areas contain multiple observations of the same albedo. Figure 11 gives some examples of material edited sequences and Figure 12 is an example of combined material editing and relighting.

7.2 Comparative Evaluation

In this section our approach is compared with segmentation and de-lighting based methods to reflectance estimation [CVH07, Gra06].

De-lighting

During the performance the studio is lit with fluorescent lighting covering the entire ceiling, giving an approximation of ambient illumination. In these conditions the main variation in illumination comes from the orientation of the surface relative to the ceiling. With surfaces facing the ceiling being lit the brightest whilst those facing progressively downwards are less brightly illuminated. Also surfaces facing progressively downwards have an increasing blue cast from the reflected light of the blue screening used in the studio. We model this with a shading map $s(n_y)$ where n_y is the y component of the surface normal. We implement this as a look-up table populated with data from an image

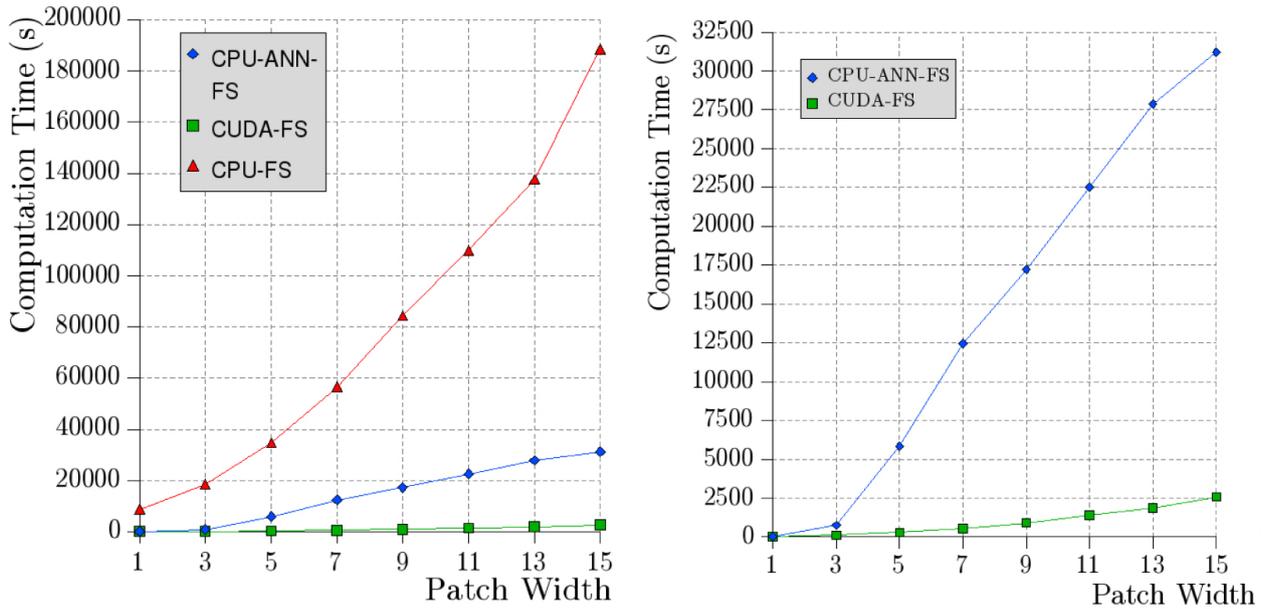


Figure 5: SSD Computation time comparison.

Patch Width	1	3	5	7	9	11	13	15
Speed-Up	0.1	6.57	20.36	22.61	18.96	16.18	14.79	12.09

Table 1: CUDA Speed-Up Relative To ANN by patch width

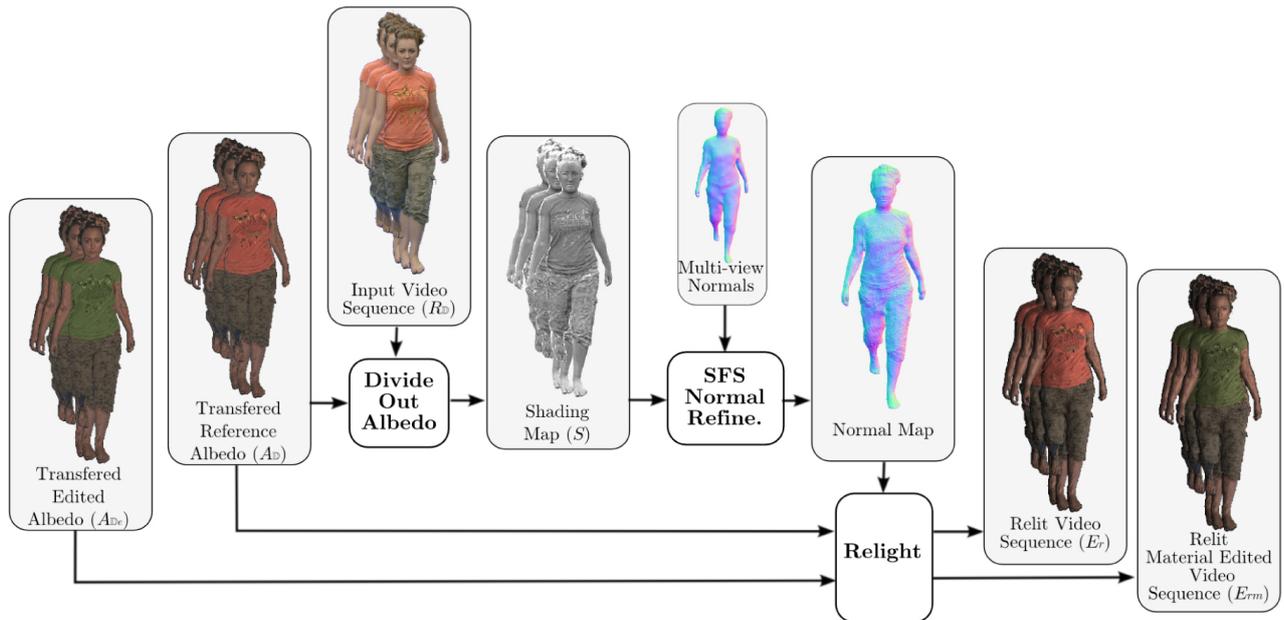


Figure 6: Combined Relighting and Material Editing Overview

of a diffuse cylinder positioned horizontally under the studio lights, giving full coverage of n_y . Then for diffuse material we can de-light using the following formula. Results for delighting are included in Figures 13 and 9.

$$D_a(i, j, t) = \frac{D_l(i, j, t)}{s(n_y(i, j, t))} \quad (14)$$

Segmentation Reflectance

For the segmentation approach we manually segmented each colour region and assigned each region a diffuse albedo as the mean colour of the region in the video frame. The original method presented in [CVH07] uses an automatic spatio-temporal segmentation allowing video to be coherently segmented, but for our evaluation on a small set of frames and a single viewpoint it was simpler to use a manual segmentation.

Evaluation

Table 2 gives a summary of the assumptions and results of the 3 relighting methods together with texture mapping. In common with the previous methods our method assumes Lambertian reflectance, but is able to separate the shape detail from the spatially varying reflectance.

Figure 13 gives a comparison of estimated normal maps, albedo maps and relit images. For the delit texture approach the large-scale shape features captured by the multi-view reconstruction are relit but the finescale details such as creases remain static throughout the different illumination conditions. The segmentation approach captures detailed shape allowing relighting of fine creases as well as the large-scale shape. However, the albedo map does not capture any of the texture details leaving these to contaminate the normal map. Our approach gives the benefits of both previous approaches with detailed reflectance variation preserved in the albedo map and fine-scale shape detail captured in the normal map. Inter-reflection and lack of illumination are potential failure points for our new approach where they are not captured in the reference set, as can be seen respectively on the actors left arm and right leg in Figure 13. The colour difference between the results for our method and the others is due to the difference in illumination sources used in the capture of the photometric stereo images and the other

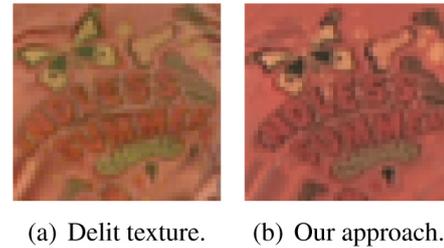


Figure 9: Texture detail albedo comparison. (a) Albedo map contaminated with shading from creases. (b) Most creases removed.

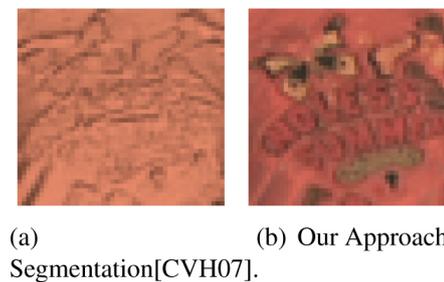


Figure 10: Texture detail relit comparison. (a) Shape detail contaminated by texture. (b) Correctly relit.

images. This could be calibrated out in the reference set.

Figure 9 presents an enlarged example of the albedo estimates. This shows how our approach is able to estimate albedo with reduced contamination from shape data whilst correctly capturing fine texture detail. Figure 10 shows a relit example for a textured area. It shows how the segmentation approach fails for detailed texture whilst our method is able to handle both the fine shape and texture detail.

8 Conclusions

This paper has introduced an approach to reflectance estimation for dynamic scenes based on transfer from a reference image set.

Local image statistics match the observed appearance with reference images captured under the same illumination conditions. This correspondence is used to transfer the known reflectance properties for the reference image to video frames of a dynamic scene.

The approach allows reflectance estimation for dynamic scenes where accurate reconstruction of the



(a) Original Video



(b) SSD, Patch Width 15



(c) SSD, Patch Width 15 and Centre Weighted



(d) NYSSD, Patch Width 15 and Centre Weighted

Figure 7: Relit Patch Match Function Comparison. The SSD examples display temporal variation due to illumination changes on the surface, whilst the NYSSD example is robust against these variations.

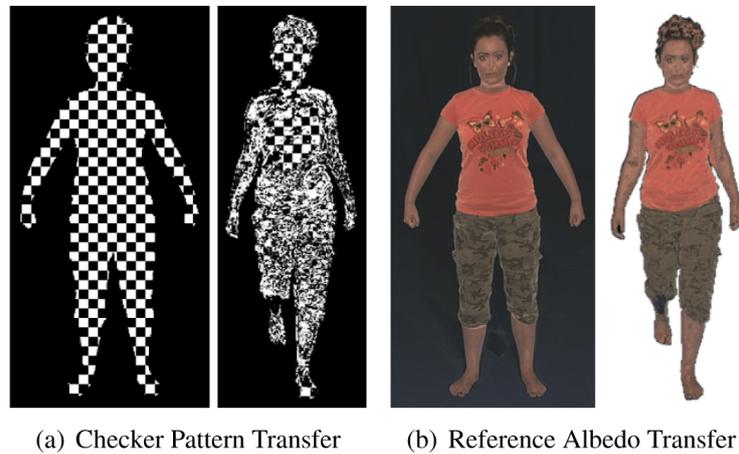


Figure 8: Reflectance Transfer Examples. (a) The transfer is only spatially consistent in areas of distinctive texture. (b) Despite spatial inconsistency the correct albedo values are transferred.

Method	Texture [SH07]	Delit Texture [Gra06]	Segmentation [CVH07]	Ours
Assumptions				
Known Illumination	×	✓	×	×
Uniform Albedo Regions	×	×	✓	×
Lambertian Reflectance	×	✓	✓	✓
Results				
Separation of Reflectance from Shape	×	×	×	✓
Large-scale Relighting	×	✓	✓	✓
Fine-scale Relighting	×	×	✓	✓

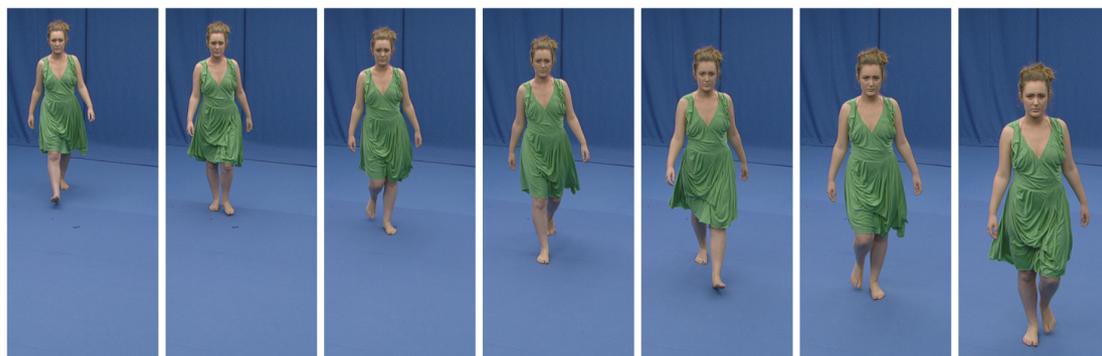
Table 2: Comparison of dynamic relighting approaches.



(a) Original



(b) Material Edited



(c) Original



(d) Material Edited

Figure 11: Material Edited Examples.

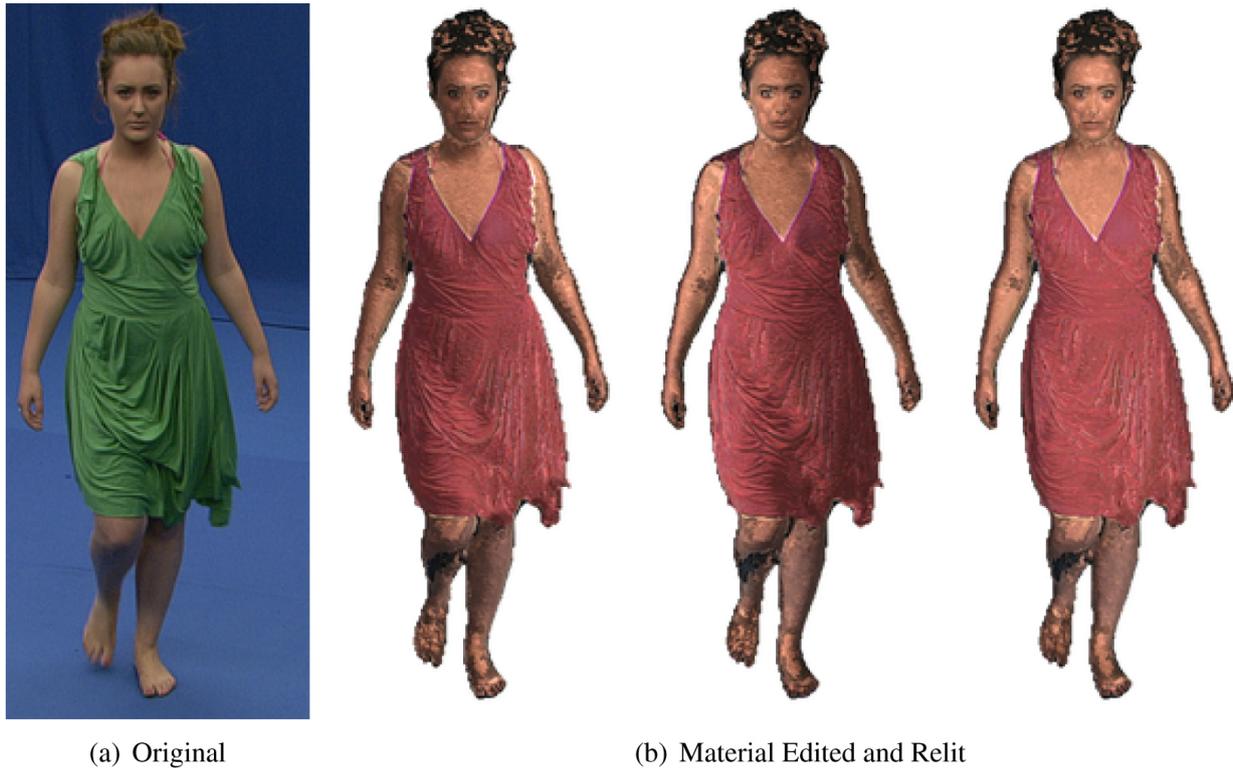


Figure 12: Material Edited and Relit Example.

shape and surface orientation is not possible. This allows transfer of reflectance and editing of material properties based on the reference image set.

Previous methods for relighting dynamic sequences have either been limited to uniform material properties or have not captured fine shading details such as those caused by creases in clothing. Reflectance transfer using local image statistics enables estimation of reflectance textured materials and capturing fine shading details.

A GPU implementation of nearest neighbour search is presented as part of the reflectance transfer method and is shown to be an order of magnitude quicker than an optimised CPU implementation.

Results are presented using the proposed reflectance transfer for relight and material editing of dynamic sequences of people with loose clothing. This demonstrates that reflectance transfer enables estimation and editing of material properties for non-rigid scenes.

Currently the approach is limited in a number of respects. Firstly, material reflectance is assumed to be Lambertian in both the photometric stereo reconstruction of reflectance properties and the material transfer. This could be addressed by a more general re-

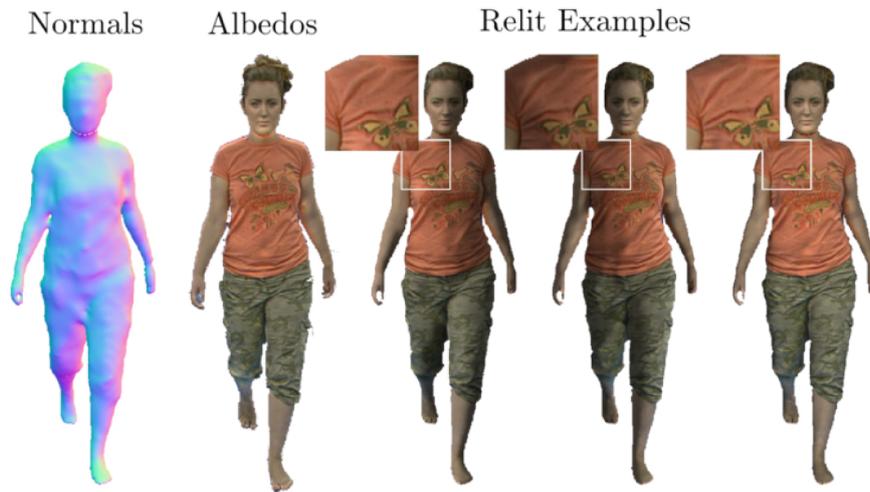
fectance reconstruction in the reference set. Secondly, the matching is performed independently on a perpixel basis and does not enforce either spatial or temporal coherence. In practice independent pixel transfer achieves a high-level of spatial coherence but results in temporal flicker. Future research will investigate temporally coherent reflectance transfer and extension of the approach to more general reflectance models.

Acknowledgements

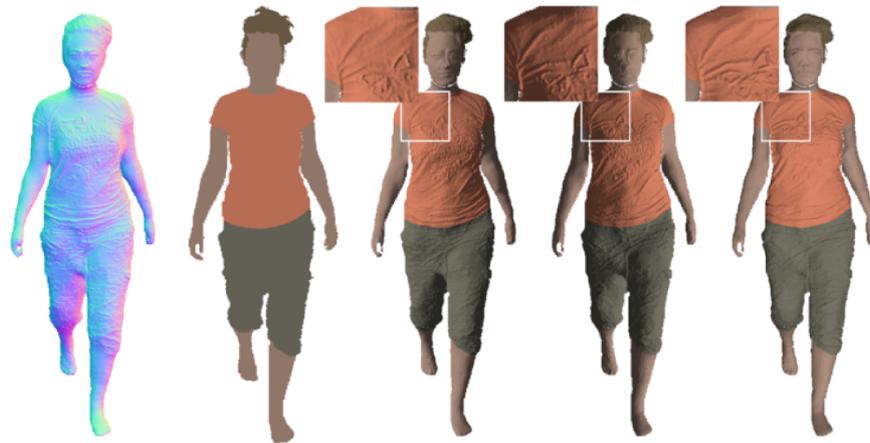
We would like to thank the EPSRC and the BBC who have supported this work through a doctoral training grant and studentship.

References

- [AMN⁺98] Sunil Arya, David M. Mount, Nathan S. Neenyahu, Ruth Silverman, and Angela Y. Wu, *An optimal algorithm for approximate nearest neighbor searching fixed dimensions*, *Journal of the ACM* **45** (1998), no. 6, 891–923, ISSN 0004-5411.



(a) Delit texture with normals from multi-view reconstruction. Similar to [Gra06].



(b) Segmentation based reflectance with SFS refined normals [CVH07].



(c) Our Approach.

Figure 13: Comparison of reflectance estimation methods. The first column shows the estimated normal maps, the second shows the acquired albedo maps and the remaining columns show relit results under different illumination conditions. (a) Creases do not respond to changing light position. (b) Dynamic lighting of creases but texture contaminates shape data instead of being captured in reflectance map. (c) Dynamic lighting of creases with texture detail separated from shape detail.

- [BP03] Svetlana Barsky and Maria Petrou, *The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows*, IEEE transactions on Pattern Analysis and Machine Intelligence **25** (2003), no. 10, 1239–1252, ISSN 0162-8828.
- [CEJ⁺06] Chales-Félix Chabert, Per Einarsson, Andrew Jones, Bruce Lamond, Wan-Chun Ma, Sebastian Sylwan, Tim Hawkins, and Paul Debevec, *Relighting Human Locomotion with flowed reflectance fields*, Rendering Techniques 2006: 17th Eurographics Workshop on Rendering, 2006, pp. 183–194.
- [CH06] Peter Csakany and Adrian Hilton, *Relighting of Facial Images*, FG 2006 7th International Conference on Automatic Face and Gesture Recognition, 2006, pp. 55–60, ISBN 0-7695-2503-2.
- [CVH07] Peter Csakany, Ferenc Vajda, and Adrian Hilton, *Recovering refined surface normals for relighting clothing in dynamic scenes*, CVMP, 2007, pp. 1–8, ISBN 978-0-86341-843-3.
- [EF01] Alexei A. Efros and William T. Freeman, *Image quilting for texture synthesis and transfer*, Computer Graphics Proceedings Siggraph 2001 (Eugene Fiume, ed.), 2001, pp. 341–346, ISBN 1-58113-374-X.
- [EL99] Alexei A. Efros and Thomas K. Leung, *Texture synthesis by non-parametric sampling*, IEEE International conference on computer vision, vol. 2, 1999, p. 1033, ISBN 0-7695-0164-8.
- [FH04] Hui Fang and John C. Hart, *Textureshop: texture synthesis as photograph editing tool*, ACM Transactions on Graphics **23** (2004), no. 3, 354–359, ISSN 0730-0301.
- [FH06] Hui Fang and John C. Hart, *Rototexture: Automated Tools for texturing raw video*, IEEE Transactions on Visualization and Computer Graphics **12** (2006), no. 6, 1580–1589, ISSN 1077-2626.
- [FWZ03] Andrew W. Fitzgibbon, Yonatan Wexler, and Andrew Zisserman, *Image-based rendering using image-based priors*, ICCV'03: Proceedings of the Ninth IEEE International Conference on Computer Vision, vol. 2, 2003, pp. 1176–1183, ISBN 0-7695-1950-4.
- [GDB08] Vincent Garcia, Eric Debreuve, and Michel Barlaud, *Fast k nearest neighbor search using gpu*, IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops 2008, CVPR, 2008, pp. 1–6, ISBN 978-1-4244-2339-2.
- [Gra06] Oliver Grau, *Multi-camera radiometric surface modelling for image-based relighting*, DAGM Symposium, Lecture Notes in Computer Science, 2006, pp. 667–676, ISBN 9783540444121.
- [HJO⁺01] A. Hertzmann, C. Jacobs, N. Oliver, B. Curless, and D. Salesin, *Image analogies*, Siggraph 2001 Conference Proceedings, 2001, pp. 327–340, ISBN 1-58113-374-X.
- [HZ07] Feng Han and Song-Chun Zhu, *A Two-level generative model for cloth representation and shape from shading*, IEEE Transactions on pattern analysis and machine intelligence **29** (2007), no. 7, 1230–1243, ISSN 0162-8828.
- [KRFB06] Erum Arif Khan, Erik Reinhard, Roland W. Fleming, and Heinrich H. Bühlhoff, *Image-based material editing*, ACM Transactions on Graphics **25** (2006), no. 3, 654–663, ISSN 0730-0301.
- [PTMD07] Pieter Peers, Naoki Tamura, Wojciech Matusik, and Paul Debevec, *Post-production facial performance relighting using reflectance transfer*, ACM Transactions on Graphics **26** (2007), no. 3, 52, ISSN 0730-0301.
- [SH07] Jonathan Starck and Adrian Hilton, *Surface capture for performance based animation*, IEEE Computer Graphics and Applications **27** (2007), no. 3, 21–31, ISSN 0272-1716.

- [SWI97] Yoichi Sato, Mark D. Wheeler, and Katsushi Ikeuchi, *Object shape and reflectance modeling from observation*, Siggraph 97: Proceedings of the 24th annual conference on computer graphics and interactive techniques, 1997, pp. 379–387, ISBN 0-89791-896-7.
- [TAL⁺07] Christian Theobald, Naveed Ahmed, Hendrik Lensch, Marcus Magnor, and Hans-Peter Seidel, *Seeing people in different light - joint shape, motion and reflectance capture*, IEEE transactions on visualization and computer graphics **13** (2007), no. 4, 663–674, ISSN 1077-2626.
- [VBS⁺07] George Vogiatzis, Gabriel J. Brostow, Björn Stenger, Roberto Cipolla, and Carlos Hernández, *Non-rigid photometric stereo with colored lights*, ICCV IEEE 11th International Conference on Computer Vision, 2007, pp. 1–8, ISBN 978-1-4244-1631-8.
- [WF05] O. Woodford and Andrew W. Fitzgibbon, *Fast image-based rendering using hierarchical image-based priors*, 16th BMVC Proceedings, 2005.

Citation
Peter Stroia-Williams, Adrian Hilton, and Oliver Grau, <i>Reflectance Transfer for Material Editing and Relighting</i> , Journal of Virtual Reality and Broadcasting, 7(2010), no. 6, September 2010, urn:nbn:de:0009-6-26537, ISSN 1860-2037.