Improved Reconstruction of Deforming Surfaces by Cancelling Ambient Occlusion

Thabo Beeler[1,2], Derek Bradley[1], Henning Zimmer[2], and Markus Gross[1,2]

[1] Disney Research Zurich
{derek.bradley,dbeeler}@disneyresearch.com
[2] ETH Zurich
{hzimmer,grossm}@inf.ethz.ch

Abstract. We present a general technique for improving space-time reconstructions of deforming surfaces, which are captured in an video-based reconstruction scenario under uniform illumination. Our approach simultaneously improves both the acquired shape as well as the tracked motion of the deforming surface. The method is based on factoring out surface shading, computed by a fast approximation to global illumination called ambient occlusion. This allows us to improve the performance of optical flow tracking that mainly relies on constancy of image features, such as intensity. While cancelling the local shading, we also optimize the surface shape to minimize the residual between the ambient occlusion of the 3D geometry and that of the image, yielding more accurate surface details in the reconstruction. Our enhancement is independent of the actual space-time reconstruction algorithm. We experimentally measure the quantitative improvements produced by our algorithm using a synthetic example of deforming skin, where ground truth shape and motion is available. We further demonstrate our enhancement on a real-world sequence of human face reconstruction.

1 Introduction

In the film and video game industries, video-based motion capture has evolved into an essential tool for generating realistic animations of actor performances. Driven by recent advances in computer vision, state-of-the-art motion capture systems can now reconstruct high-quality deforming surfaces in dense correspondence, enabling performance capture at the resolution of deforming cloth and skin [1–6]. The goal of these systems is to reconstruct both the time-varying shape as well as the motion for each point on the surface, typically utilizing tools such as multi-view 3D reconstruction [7] and image-based tracking via dense optical flow [8].

Despite the high quality reconstructions that these systems achieve, we identify two problems that can occur when reconstructing the shape and motion of a deforming surface from video images. First, as a surface deforms it exhibits local shading changes over time, which tend to decrease the accuracy of most optical flow tracking methods, as they typically assume constant brightness (or constant

A. Fitzgibbon et al. (Eds.): ECCV 2012, Part I, LNCS 7572, pp. 30–43, 2012.

brightness gradients) of corresponding pixels. To minimize shading changes from shadows and specular highlights, capture setups often take special care to capture surfaces under diffuse ambient illumination. Even then, local shading changes can occur during deformation, particularly for local high-frequency changes, such as the folds of cloth or the wrinkles of skin. As a result, the temporal image tracking of these interesting phenomena is often inaccurate, leading to an incorrect motion estimation. To address this issue, we propose a technique for improving the motion reconstruction of a deforming surface under uniform illumination by increasing the accuracy of the image-space tracking. We observe that shading changes in these ambient illumination setups can be well-approximated by ambient occlusion. Therefore, we compute the per-frame ambient occlusion of the deforming surface and cancel this term from the input images, resulting in an image-sequence that almost perfectly fulfills the brightness constancy assumption used in the majority of optical flow methods. We show that optical flow tracking is more accurate on this sequence than on the original images. This general improvement is orthogonal to the optical flow implementation, and we demonstrate that it complements several well known optical flow algorithms.

The second problem is that the reconstructed 3D shape of local high-frequency details also tends to be inaccurate. This is often a result of poor photometric information caused by foreshortening and local shading, as well as over-smoothing of surface details to eliminate noise. We can again use ambient occlusion to correct this problem. If the initial shape is incorrect, then the estimated ambient occlusion of the surface will differ from the measured shading in the image. We use this residual to improve the 3D geometry, by optimizing for the shape that best matches the ambient occlusion observed in the images. Like our motion improvement technique, the geometry enhancement is independent of the initial reconstruction algorithm.

Our two-phase process of improving motion tracking and enhancing the reconstructed geometry can be repeated iteratively until convergence. When applied to a sequence captured under uniform illumination, the result is a more accurate spatio-temporal reconstruction of the deforming surface, which is more faithful to the true surface both in terms of shape as well as motion.

1.1 Related Work

We will first review motion tracking under brightness changes, and then discuss methods related to our illumination-aware geometry enhancement technique.

Illumination-Robust Optical Flow. There are several strategies for dealing with brightness changes in optical flow estimation. For purely additive changes imposing constancy of the image gradient is one solution, used for example in [9]. More sophisticated methods consider patch-based matching scores like cross correlation [10] that are more invariant, e.g. under multiplicative changes. However, higher invariance always discards image information that could help the matching in regions without brightness changes and implementing sophisticated matching scores can be very tedious. Even more involved are approaches that

estimate spatially varying maps that explain multiplicative [11] or additive and multiplicative brightness changes, e.g. [12]. This introduces additional unknowns that need to be estimated which further complicates implementation and introduces more tuning parameters that need to be carefully adjusted. Similarly, Haussecker and Fleet [13] jointly estimate optical flow and model parameters of a physical processes causing brightness change. However, the processes are restricted to a moving illumination source and surface rotation. If color images are available, using alternative color spaces like the HSV space, can also improve robustness under illumination changes; see [14] and the references therein. However, such color spaces also cause problems, e.g. the H- and S-channel of the HSV space cannot distinguish grayscales [14]. In practice, we also found no color space transformation that can adequately separate illumination changes from albedo, e.g. when skin wrinkles occur.

Similar to our approach, Wedel et al. [15] also modify the input images to ease flow estimation. They decompose the images into a structure and a texture part showing large scale image features and high frequency details, respectively. Then, mainly the texture part is used for flow computation as illumination artifacts are mainly present in the structure part. However, this only holds for larger objects casting shadows and not for small details like wrinkles, which are in the focus of our work. Moreover, the texture part also contains most of the noise and thus decreases the robustness of the flow estimation. Related to this are retinex techniques like homomorphic filtering [16] that separate the image in reflectance and illumination components, which is however an inherently ill-posed problem.

In summary we can say that in contrast to previous work, our strategy leverages information about the 3D scene geometry to cancel out brightness changes. In this, we certainly tackle less general scenarios, but being tailored towards a specific application in mind, i.e. the temporal tracking of deforming surfaces, we achieve high accuracy results even using readily available optical flow algorithms.

Illumination-Aware Geometry Reconstruction. Reconstructing 3D shape from surface shading is a well studied area of computer vision. Since the early work of Horn on shape-from-shading [17] and Woodham on photometric stereo [18], many variants of this approach have emerged, including a lot of work that fuses shading cues with multi-view information (just a few recent examples are [19–21]). Related to our geometry enhancement technique, the shape reconstruction method of Wu et al. [22] performs shading-based shape refinement for smooth surfaces reconstructed from multi-view stereo. They estimate a general illumination model of the scene and refine the surface such that the shading variations observed in the images are explained by the geometry. However, they assume the surface has a constant albedo, which limits the application range. For example, it would not be feasible to refine real human face reconstructions without producing artifacts.

For deformable object reconstruction, it is possible to obtain highly detailed geometry under calibrated lighting [23–25], however this places a heavy restriction on the capture setup. We target lighting setups that provide simple uniform illumination and we use ambient occlusion to refine an initial surface estimate

obtained from multi-view stereo. Illumination-based refinement of deforming surface reconstructions has been studied previously. Popa et al. [26] add wrinkles to deforming garments by identifying the shading edges of cloth folds in the video images. The geometry is then refined to synthesize wrinkles in a temporally consistent manner. Wu et al. also propose to correct wrinkles on cloth reconstructions [27]. Building on their previous work [22], they estimate the environment lighting and surface albedo, then add missing fine-scale details to a coarse dynamic mesh in a temporally consistent way. Both of these techniques assume that the temporal motion is accurate from the start, whereas we couple geometry and motion refinement into an iterative scheme for globally improving the spatio-temporal reconstruction.

1.2 Contributions

In this paper we propose techniques to simultaneously improve the time-varying shape as well as the motion tracking of a deforming surface, captured in a video-based reconstruction scenario under uniform illumination. Specifically, our contributions are as follows:

1. We show that ambient occlusion can be computed on the per-frame geometry and removed from the input images, improving the performance of several well-known optical flow algorithms.
2. We propose a shape optimization method that minimizes the ambient occlusion residual between the geometry and the images.
3. We evaluate our method on ground truth data and highlight the practicality of our approach by enhancing a real-world sequence of face deformation.

2 Problem Definition and Method Overview

To reconstruct a deforming surface, we must capture both the time-varying shape of the surface as well as its motion, tracking each point over time. In recent approaches, this has been achieved by combining multi-view reconstruction techniques with image-based optical flow tracking [1, 5, 6]. As we mentioned in the introduction, if the surface contains local high-frequency deformations then both the optical flow tracking and the reconstructed surface shape can be inaccurate. Figure 1 highlights these problems for a real-world reconstruction example of deforming skin [6]. Here we see that the local shading variation causes incorrect flow vectors for a well-known optical flow technique [9]. Furthermore, the reconstructed surface geometry [28] is overly smooth in the wrinkle regions. In order to compensate for these problems, our approach is to compute and factor out local surface shading, approximated by ambient occlusion.

2.1 Ambient Occlusion

Ambient occlusion is a global shading method that approximates global illumination [29]. It does not take into account effects such as cast shadows,

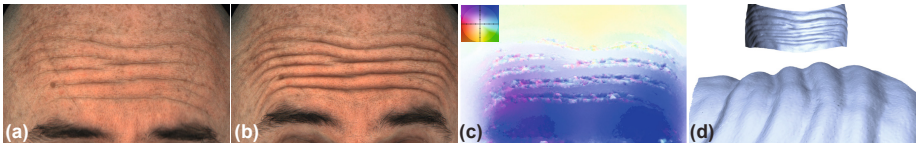

Fig. 1. Motion and geometry reconstruction problems for deforming skin. Two video images (a) and (b) showing a skin wrinkle forming. (c) shows that a state-of-the-art optical flow algorithm [9] fails for such illumination changes (flow vectors are visualized using the color code shown in the inset). (d) shows that the reconstructed geometry [28] is overly smooth.

inter-reflections or subsurface scattering. However, in a setting with diffuse or omnidirectional illumination, ambient occlusion approximates global illumination well. Ambient occlusion is defined as

$$A(\mathbf{x}) = \frac{1}{\pi} \int_{\Omega} V(\mathbf{x}, \omega) \langle \mathbf{n}(\mathbf{x}), \omega \rangle \, d\omega, \tag{1}$$

where $\mathbf{x}$ is a point on the surface, $\mathbf{n}(\mathbf{x})$ the normal at this point, $\langle \cdot, \cdot \rangle$ denotes the inner product and $V(\cdot, \cdot)$ is a visibility function that is 0 if the ray ω is occluded and 1 otherwise. The integral is formed over the hemisphere Ω, which makes ambient occlusion costly to compute in general, especially for large meshes. Several methods have been proposed for efficient ambient occlusion approximations (see [30] for a survey).

In our work, we will refine the shape of a surface based on the computed ambient occlusion, and so the quality of the refinement depends on the accuracy of the ambient occlusion estimation. At the same time, our refinement method is iterative, and so we aim for fast computation. To meet these requirements, we implemented a fast ray-tracing approach with deterministic ray samples. We use the Intel Embree high-performance ray-tracing library[1], which is designed exactly for this purpose. However, instead of traditional Monte-Carlo stochastic ray tracing, we use a deterministic cosine-distribution of samples around each vertex normal. While this approach introduces a small amount of bias in the result, the spatially-varying noise in ambient occlusion is greatly reduced for the same number of samples, allowing us to compute a close approximation in a matter of seconds rather than tens of minutes with the Monte-Carlo approach. Figure 2 shows the ambient occlusion computation for a 3D surface patch. In order to determine how many ray samples to use, we plot the RMS error for different sample sizes (Figure 2 (d)) compared to ground truth ambient occlusion computed using Monte-Carlo ray-tracing with 100,000 samples (Figure 2 (b)). We found that any quality gain beyond 500 samples (Figure 2 (c)) was negligible. Note that our deterministic approximation does not converge exactly to the ground truth, as we see in the RMS error plot, and the convergence is not a monotonically decreasing function. However, the result is visually almost

[1] `http://software.intel.com/en-us/articles/embree-photo-realistic-ray-tracing-kernels/`

identical to the ground truth, and in practice we found that inaccuracies of such a small magnitude had no effect on our algorithm.

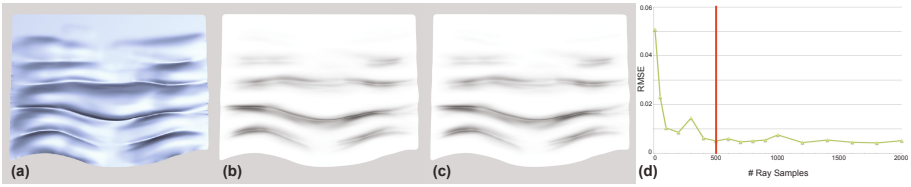

Fig. 2. Ambient occlusion computation. (a) 3D surface patch. (b) Ground-truth computed using Monte-Carlo ray-tracing with 100,000 ray samples. (c) Our approximation with 500 samples is almost identical. (d) RMS error of our approximation for different sample sizes.

2.2 Method Overview

We now give an overview of our technique for improving reconstructions of deforming surfaces by cancelling ambient occlusion. A pictorial representation of the method can be found in Figure 3. Given a sequence of reconstructed meshes and the corresponding calibrated camera images, our algorithm processes the frames sequentially with three main steps per frame:

1. **Cancel Ambient Occlusion** - The ambient occlusion of the surface is computed and projected onto each image plane, then divided out of the image. Ambient occlusion is computed as described in Section 2.1.
2. **Motion Improvement** - Optical flow is computed on the shading-free images created by cancelling ambient occlusion (Section 3).
3. **Shape Improvement** - The 3D shape is refined to minimize the ambient occlusion residual when cancelling from the images, leading to a surface that better captures fine details such as wrinkles (Section 4).

These steps are iterated until the shape and motion refinement becomes negligible. In our experiments, typically only two to three iterations are required.

Our algorithm is completely independent of the original 3D reconstruction method and the optical flow algorithm. In Section 5 we will show improvements to the state-of-the-art facial geometry reconstruction method [28]. We also show that several common optical flow methods [31, 32, 9, 14, 33] perform much better on the ambient occlusion cancelled images than the original images.

2.3 Notation

The following notation will be used throughout the paper.
$\mathcal{M}^t(\mathbf{x})$ is a mesh at frame t, defined over vertices $\mathbf{x}$. $\mathcal{M}^t$ will be used for short.
$\mathcal{I}^t(\mathbf{p})$ is an image at frame t, defined over pixels $\mathbf{p}$. $\mathcal{I}^t$ will be used for short.
$\mathcal{F}^{t-1\leftarrow t}(\mathbf{p})$ is the flow from $\mathcal{I}^t$ to $\mathcal{I}^{t-1}$. $\mathcal{F}^{t-1\leftarrow t}$ will be used for short.

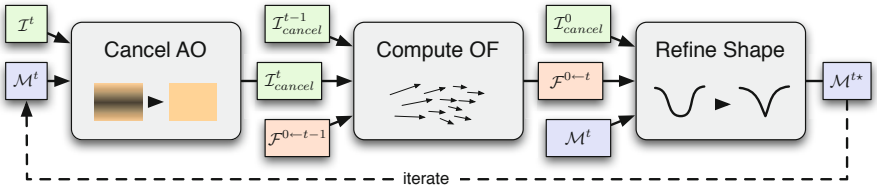

Fig. 3. Overview of the algorithm - the algorithm consists of three major stages. Stage 1 computes ambient occlusion on the mesh $\mathcal{M}^t$ and removes it from the input image $\mathcal{I}^t$. This improves the estimation of the flow field $\mathcal{F}^{0\leftarrow t}$ in Stage 2. Stage 3 produces a refined shape $\mathcal{M}^{t\star}$ by minimizing the residual of the observed shading and computed ambient occlusion.

3 Motion Improvement

The flow field $\mathcal{F}^{t-1\leftarrow t}$ is improved by removing shading caused by ambient occlusion of the reconstructed meshes $\mathcal{M}^{t-1}$ and $\mathcal{M}^t$ from the images $\mathcal{I}^{t-1}$ and $\mathcal{I}^t$, respectively. The shading is removed via

$$\mathcal{I}^t_{cancel} = \frac{\mathcal{I}^t}{P\left(A(\mathcal{M}^t)\right)}, \tag{2}$$

where $P(\cdot)$ projects the ambient occlusion computed for the mesh $\mathcal{M}$ onto the image $\mathcal{I}$. The improved flow field $\mathcal{F}^{t-1\leftarrow t}_{cancel}$ is integrated with $\mathcal{F}^{0\leftarrow t-1}_{cancel}$ to produce the motion estimation $\mathcal{F}^{0\leftarrow t}_{cancel}$ from frame t to the first frame. The flow is estimated backwards to facilitate easy warping of the first frame to frame t, which will be required in the next stage of the algorithm.

4 Shape Improvement

The motivation to refine the shape stems from the observation that in the assumed setting most shading changes are caused by shape deformation. The shape is refined such that the predicted shading corresponds to the observed shading. The shape improvement consists of the following steps:

1. Compute the observed shading $A'(\mathbf{x})$ from the images.
2. Compute the ambient occlusion $A(\mathbf{x})$ on the surface.
3. Compute the refinement $\delta(\mathbf{x})$ based on $A'(\mathbf{x})$ and $A(\mathbf{x})$.
4. Update vertex positions $\mathbf{x}^\star = \mathbf{x} + \delta(\mathbf{x})\mathbf{n}(\mathbf{x})$.

These steps are performed iteratively for all vertices of a mesh. Note that vertices are displaced only along the normal direction. Constraining the refinement to a single dimension greatly reduces computational complexity and increases robustness of the algorithm. If the surface contains many high-frequency details, we found that a low-pass filter of the normals produces better results. Although

normal vectors are updated in each iteration in order to compute accurate ambient occlusion, the displacement directions of the vertices remain constant. The steps of our algorithm are explained in more detail in the following.

The observed shading is computed from the input images. For a single image, $A'(\mathbf{x})$ is computed as

$$A'(\mathbf{x}) = \frac{I^t(\mathbf{q})}{W^{0\to t}(I^0_{cancel}(\mathbf{q}))}, \tag{3}$$

where $\mathbf{q}$ is the projection of $\mathbf{x}$ onto the image plane, and $W^{a\to b}$ is a warping function that warps an image from frame a to frame b given the flow field $\mathcal{F}^{a\leftarrow b}_{cancel}$ computed in Section 3. If multiple views exist, $A'(\mathbf{x})$ can be computed as a (weighted) average from all images. In this equation we assume that shading has been completely removed from I^0_{cancel}, which is plausible if I^0 is chosen as a neutral expression.

The predicted shading $A(\mathbf{x})$ is computed from the mesh $\mathcal{M}^t$ using ambient occlusion, as described in Section 2.1. The refined position $\mathbf{x}^\star$ for a vertex $\mathbf{x}$ of the mesh $\mathcal{M}^t$ is computed as

$$\mathbf{x}^\star = \mathbf{x} + \delta(\mathbf{x})\mathbf{n}(\mathbf{x}). \tag{4}$$

Using the residual $\delta_A(\mathbf{x}) = s(A(\mathbf{x}) - A'(\mathbf{x}))$, where s converts the unit-less ambient occlusion to an appropriate scale for the geometry, the refinement is computed as

$$\delta(\mathbf{x}) = \frac{\gamma(A'(\mathbf{x}))\delta_A(\mathbf{x}) + \lambda\delta_L(\mathbf{x})}{\gamma(A'(\mathbf{x})) + \lambda}, \tag{5}$$

where λ is a parameter that controls the influence of the regularization. We use $\lambda = 2$. The regularized offset $\delta_L(\cdot)$ is computed using Laplacian coordinates as

$$\delta_L(\mathbf{x}) = \langle \nabla^2\mathcal{M}(\mathbf{x}), \mathbf{n}(\mathbf{x})\rangle - \eta(\mathbf{x}), \tag{6}$$

where $\eta(\cdot)$ controls the target shape. The default choice is $\eta(\mathbf{x}) = 0$ for all vertices, which prefers smooth solutions. If the shape of the input meshes can be considered mostly accurate, then setting $\eta(\cdot)$ to the Laplacian coordinates of the input mesh is a better choice. In these cases the regularization will try to maintain the input shape.

The non-linear function $\gamma(\cdot)$ controls the refinement strength depending on the observed shading $A'(\mathbf{x})$. This function accounts for the non-linear influence of noise in $A'(\mathbf{x})$ on the shape. The same perturbation of $A'(\mathbf{x})$ would induce larger perturbation of the shape in areas of lower concavity, as depicted in Figure 4.

From the illustration in Figure 4, an ambient occlusion value can be characterized by a half-angle α, defining a cone of visibility. Then $\gamma(\cdot)$ can be written as a function of α,

$$\gamma(\alpha) = \frac{\cos(\alpha) + \epsilon}{1 + \epsilon}, \tag{7}$$

where ϵ is a small parameter that controls the lower bound of $\gamma(\cdot)$. Setting ϵ to 0 will prevent refinement in planar and convex areas. We use $\epsilon = 0.1$. The observed shading $A'(\mathbf{x})$ is related to $\gamma(\cdot)$ via the angle α through

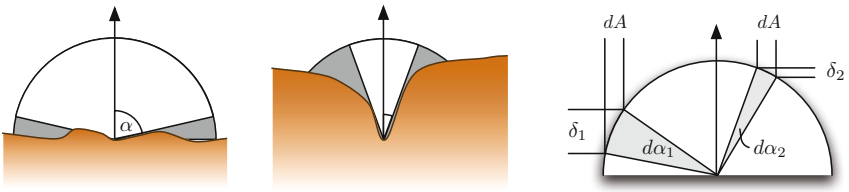

Fig. 4. This figure depicts the non-linear dependency of the displacement on the observed shading in 2D. The left and middle drawings show the relation of the half-angle α to the ambient occlusion for two different cases. The right figure illustrates that the same perturbation dA in ambient occlusion leads to different $d\alpha$ and therefore different displacements δ depending on the concavity of the surface.

$$A'(\mathbf{x}) \approx \frac{1}{\pi} \int_{\Omega} V(\mathbf{x}, \omega) \langle \mathbf{n}(\mathbf{x}), \omega \rangle \, d\omega \tag{8}$$

$$= \frac{1}{\pi} \int_0^{2\pi} \int_0^{\alpha} \cos(\phi) \sin(\phi) \, d\phi d\theta \tag{9}$$

$$= \sin^2(\alpha) \tag{10}$$

and thus

$$\gamma(A'(\mathbf{x})) = \frac{\sqrt{1 - A'(\mathbf{x})} + \epsilon}{1 + \epsilon}. \tag{11}$$

Refining convex areas less than concave ones is also very valuable when assuming the input shape is mostly accurate, as areas that are well visible are more likely to be of correct shape than concave ones.

5 Results

In this section we present the results of our algorithm, starting with an evaluation using a synthetic dataset, and then demonstrating the improvement we achieve on a real-world capture sequence.

5.1 Quantitative Evaluation

To quantitatively measure the effect of our optical flow improvement method, we designed a synthetic test sequence giving us ground truth motion. The sequence consists of a skin-textured surface patch that undergoes wrinkling while deforming. The deformation was created in Maya using blend shapes. We evaluate six different optical flow algorithms: pyramid implementations of Lucas-Kanade [31] and Horn-Schunck [32], as well as the Horn-Schunck algorithm with added gradient constancy, the method of Brox et al. [9], Zimmer et al. [14], and Werlberger

et al. [33]. We used the openCV[2] implementation for Lucas-Kanade, the flowLib[3] library for Werlberger et al. and the implementations of the other algorithms were kindly provided by Zimmer et al. [14]. Figure 5 shows the two frames that were used for the benchmark, as well as the backward flow field. The flow visualization from the Middlebury evaluation [8] is not very meaningful in this case (Figure 5 (c)) since the motion is primarily 1-dimensional, so we use a custom visualization based on the flow magnitude for this evaluation (Figure 5 (d)).

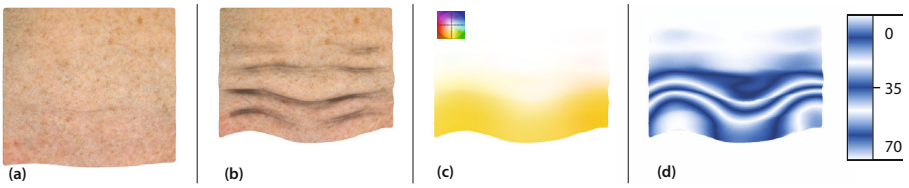

Fig. 5. Figures (a) and (b) show two frames of the synthetic sequence used in the quantitative evaluation. (c) shows the ground truth flow field in the Middlebury color scheme. To better visualize the large variation in displacement (0px-70px) we employ the iso-contour scheme depicted in (d).

We chose two temporally distant frames on purpose, to better visualize the impact of the method. The flow algorithms are also improved for closer frames but the effect is, of course, smaller if there is less change in shading. A selection of the evaluation benchmark results are shown in Figure 6 and the full results are listed in Table 1. We report the Endpoint Error [8] for the complete patch as well as the Endpoint Error for the worst 33% in order to account for the fact that only part of the image exhibits shading change.

Table 1. Mean Endpoint Errors (EE) and standard deviations for all benchmarked algorithms. The proposed method greatly improves the performance of all algorithms. The errors reported for Lucas and Kanade are less indicative because they include outliers that are not caused by wrinkling and the completeness of the result differ substantially (25% on the original and 87% on the cancelled sequence).

Algorithm	Endpoint Error (EE) [px]		Upper 33% EE [px]	
	Original	Refined	Original	Refined
Horn-Schunck	13.86±12.06	0.15±0.16	28.79±5.85	0.31±0.19
Horn-Schunck (grad.)	3.27±7.83	0.15±0.18	9.86±11.34	0.32±0.22
Brox et al.	1.77±6.03	0.13±0.14	10.72±22.22	0.29±0.16
Zimmer et al.	2.53±6.89	0.15±0.20	8.38±12.21	0.32±0.28
Werlberger et al.	3.14±7.07	0.33±1.14	9.55±9.82	0.76±1.95
Lucas-Kanade	3.02±8.34	1.38±5.59	278.27±85.73	24.12±52.84

[2] `http://sourceforge.net/projects/opencvlibrary`
[3] `http://gpu4vision.icg.tugraz.at/index.php?content=subsites/flowlib/flowlib.php`

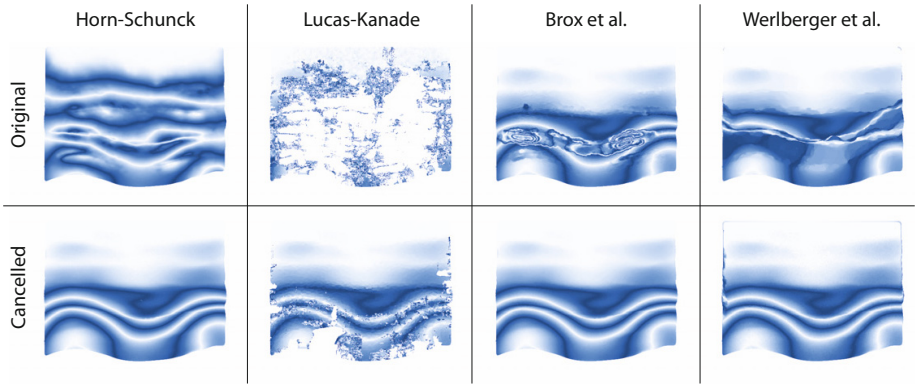

Fig. 6. This figure shows computed flow fields on the original (top row) and cancelled images (bottom row) for four well known algorithms. All algorithms have problems in areas where their core matching assumptions are violated due to the change in shading. Cancelling ambient occlusion increases the performance of all algorithms substantially. The ground truth flow field is shown in Figure 5 and the computed errors are listed in Table 1.

As can be seen from Figure 6, the flow methods have problems estimating the motion within the wrinkles due to shading change. Cancelling the ambient occlusion gives consistently better results for all algorithms. Algorithms that assume only brightness constancy benefit most, but also algorithms that include gradient constancy show better performance. On the cancelled sequence, the gradient constancy assumption is not that influential anymore and, for example, Brox et al. produce the lowest Endpoint Error without gradient constancy. A special case is Lucas-Kanade. While its performance is greatly improved (see the second column of Figure 6) the reported errors are less indicative because they include outliers that are not caused by wrinkling and the completeness of the result differs substantially (25% on the original and 87% on the cancelled sequence).

The parameters were empirically chosen to produce the best result that can be achieved before and after cancelling ambient occlusion. In general, we found that parameter tuning for the original images was more challenging since the resulting flow was very sensitive to parameter changes, where on the cancelled images flow computation was more robust to parameter variations. To achieve decent flow estimations for the original images we had to choose high regularization parameters, while for the cancelled images significantly lower regularization produced the most accurate results.

5.2 Real-World Sequence

Figure 7 demonstrates the impact of the method on a real-world sequence (kindly provided by Beeler et al. [6]). Motivated by the analysis of the synthetic data (Table 1) we employed the method of Brox et al. to compute the flow as it

performed best with similar parameter settings for the original and cancelled images. We ran three overall iterations and the mesh refinement ran for 30 iterations. The runtime of our parallelized C++ implementation was under 3 minutes per frame on a Mac Pro using 8 cores. The size of the images is 539×329 pixels and the mesh consists of roughly 150K vertices. As can be seen, not only the flow field but also the shape of the wrinkles is greatly improved. Figure 8 shows that the wrinkles of the original and refined meshes produce substantially different silhouettes.

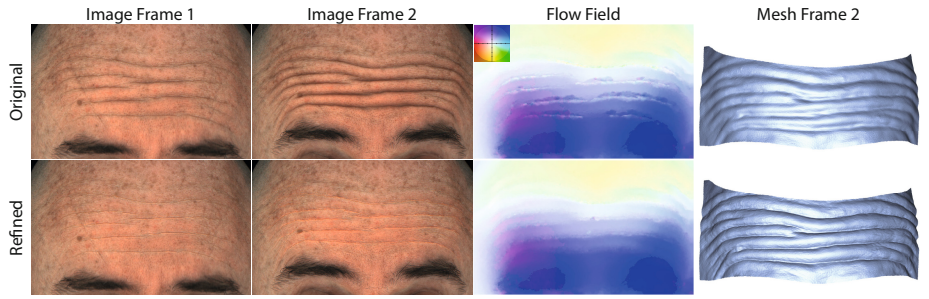

Fig. 7. Real-World Sequence - the first row shows the original images along with computed flow field and input shape. The second row shows the cancelled images along with the improved flow field and refined shape after three iterations. The flow fields were computed using the method of Brox et al.

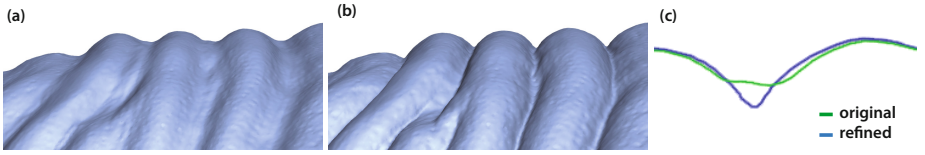

Fig. 8. Shape Refinement - (a) and (b) show the original and refined shapes. Figure (c) overlays the silhouette of both shapes for better comparison. Note how the refined shape exhibits the v-shaped valleys and u-shaped ridges characteristic to wrinkles, while the original shape fails to do so.

6 Discussion and Conclusion

We have shown that cancelling the ambient occlusion from captured images of a deforming surface can lead to improved motion and surface reconstructions, particularly for high-frequency deformations such as the wrinkles of human skin.

In order to accurately improve reconstructions, our method makes a few assumptions about the captured sequence. First, we assume that the lighting setup is very close to ambient illumination, so that the recorded shading can be closely approximated by ambient occlusion. This is not a severe limitation however, because most capture setups are designed with nearly uniform omni-directional lighting. In the case of arbitrary non-uniform illumination, we could estimate

the environment lighting using a method similar to Wu et al. [27], which we consider future work. The second assumption is that the reconstructed sequence contains at least one frame without any shading, or alternatively the correct shape such that ambient occlusion can be computed and removed. We have referred to this frame as the reference mesh $\mathcal{M}^0$ and reference image $\mathcal{I}^0$. This frame is needed to remove albedo and compute the observed shading in the rest of the sequence (Equation 3). For face reconstructions, this can be a neutral pose without wrinkles. If no such frame exists, a simple pre-processing step could be applied to search for the brightest occurrence of each vertex in the sequence, assuming that it is un-shaded at some point. We will explore this direction in future work.

This paper includes a detailed evaluation of our technique, using a ground-truth example of a deforming surface patch with known motion. Several well-known optical flow algorithms are shown to benefit from our approach. This demonstrates that when considering a specific scenario, like tracking deforming surfaces, modifying the input data in an appropriate way can be a valuable alternative to designing more sophisticated optical flow algorithms.

We demonstrate our method on a real-world capture sequence of a human face undergoing an expression change. In particular, our refinement process produces much more realistic skin wrinkles. Our method can be applied to an existing reconstruction sequence, independently of the reconstruction method or the optical flow algorithm, and it can be easily integrated into new space-time reconstruction algorithms.

References

1. de Aguiar, E., Theobalt, C., Stoll, C., Seidel, H.-P.: Marker-less deformable mesh tracking for human shape and motion capture. In: CVPR (2007)
2. Bradley, D., Popa, T., Sheffer, A., Heidrich, W., Boubekeur, T.: Markerless garment capture. ACM Trans. Graphics (Proc. SIGGRAPH), 99 (2008)
3. de Aguiar, E., Stoll, C., Theobalt, C., Ahmed, N., Seidel, H.P., Thrun, S.: Performance capture from sparse multi-view video. ACM Trans. Graphics (Proc. SIGGRAPH), 98 (2008)
4. Vlasic, D., Baran, I., Matusik, W., Popović, J.: Articulated mesh animation from multi-view silhouettes. ACM Trans. Graphics (Proc. SIGGRAPH), 97 (2008)
5. Bradley, D., Heidrich, W., Popa, T., Sheffer, A.: High resolution passive facial performance capture. ACM Trans. Graphics (Proc. SIGGRAPH) (2010)
6. Beeler, T., Hahn, F., Bradley, D., Bickel, B., Beardsley, P., Gotsman, C., Sumner, R.W., Gross, M.: High-quality passive facial performance capture using anchor frames. ACM Trans. Graphics (Proc. SIGGRAPH) 30, 75 (2011)
7. Seitz, S.M., Curless, B., Diebel, J., Scharstein, D., Szeliski, R.: A comparison and evaluation of multi-view stereo reconstruction algorithms. In: CVPR (2006)
8. Baker, S., Scharstein, D., Lewis, J.P., Roth, S., Black, M.J., Szeliski, R.: A database and evaluation methodology for optical flow. IJCV 92, 1–31 (2011)
9. Brox, T., Bruhn, A., Papenberg, N., Weickert, J.: High Accuracy Optical Flow Estimation Based on a Theory for Warping. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3024, pp. 25–36. Springer, Heidelberg (2004)
10. Molnár, J., Chetverikov, D., Fazekas, S.: Illumination-robust variational optical flow using cross-correlation. CVIU 114, 1104–1114 (2010)

11. Seitz, S.M., Baker, S.: Filter flow. In: ICCV (2009)
12. Gennert, M., Negahdaripour, S.: Relaxing the brightness constancy assumption in computing optical flow. Technical Report A.I. Memo No. 975. MIT (1987)
13. Haussecker, H.W., Fleet, D.J.: Computing optical flow with physical models of brightness variation. TPAMI 23, 661–673 (2001)
14. Zimmer, H., Bruhn, A., Weickert, J.: Optic flow in harmony. IJCV 93(3), 368–388 (2011)
15. Wedel, A., Pock, T., Zach, C., Bischof, H., Cremers, D.: An Improved Algorithm for TV-L^1 Optical Flow. In: Cremers, D., Rosenhahn, B., Yuille, A.L., Schmidt, F.R. (eds.) Statistical and Geometrical Approaches to Visual Motion Analysis. LNCS, vol. 5604, pp. 23–45. Springer, Heidelberg (2009)
16. Stockham, T.G.: Image processing in the context of a visual model. Proc. IEEE 60 (1972)
17. Horn, B.K.P.: Shape from Shading: A Method for Obtaining the Shape of a Smooth Opaque Object from One View. PhD thesis. MIT (1970)
18. Woodham, R.J.: Photometric method for determining surface orientation from multiple images. Optical Engineering 19(1), 139–144 (1980)
19. Joshi, N., Kriegman, D.: Shape from varying illumination and viewpoint. In: ICCV (2007)
20. Hernandez, C., Vogiatzis, G., Cipolla, R.: Multiview photometric stereo. TPAMI 30 (2008)
21. Wu, C., Liu, Y., Dai, Q., Wilburn, B.: Fusing multiview and photometric stereo for 3d reconstruction under uncalibrated illumination. TVCG 17, 1082–1095 (2011)
22. Wu, C., Wilburn, B., Matsushita, Y., Theobalt, C.: High-quality shape from multi-view stereo and shading under general illumination. In: CVPR (2011)
23. Hernandez, C., Vogiatzis, G., Brostow, G.J., Stenger, B., Cipolla, R.: Non-rigid photometric stereo with colored lights. In: ICCV (2007)
24. Ahmed, N., Theobalt, C., Dobrev, P., Seidel, H.-P., Thrun, S.: Robust fusion of dynamic shape and normal capture for high-quality reconstruction of time-varying geometry. In: CVPR (2008)
25. Vlasic, D., Peers, P., Baran, I., Debevec, P., Popović, J., Rusinkiewicz, S., Matusik, W.: Dynamic shape capture using multi-view photometric stereo. ACM Trans. on Graphics 28(5), 174 (2009)
26. Popa, T., Zhou, Q., Bradley, D., Kraevoy, V., Fu, H., Sheffer, A., Heidrich, W.: Wrinkling captured garments using space-time data-driven deformation. CGF (Proc. Eurographics) 28(2), 427–435 (2009)
27. Wu, C., Varanasi, K., Liu, Y., Seidel, H.-P., Theobalt, C.: Shading-based dynamic shape refinement from multi-view video under general illumination. In: ICCV (2011)
28. Beeler, T., Bickel, B., Sumner, R., Beardsley, P., Gross, M.: High-quality single-shot capture of facial geometry. ACM Trans. Graphics (Proc. SIGGRAPH) (2010)
29. Zhukov, S., Iones, A., Kronin, G.: An ambient light illumination model. In: Proc. of Eurographics Workshop on Rendering, pp. 45–55 (1998)
30. Méndez-Feliu, A., Sbert, M.: From obscurances to ambient occlusion: A survey. Visual Computer 25(2), 181–196 (2009)
31. Lucas, B., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: IJCAI, pp. 674–679 (1981)
32. Horn, B.K.P., Schunck, B.G.: Determining optical flow. Artificial Intelligence 17, 185–203 (1981)
33. Werlberger, M., Pock, T., Bischof, H.: Motion estimation with non-local total variation regularization. In: CVPR (2010)