From Differential Photometric Consistency to Surface Differential Geometry

Anonymous CVPR submission

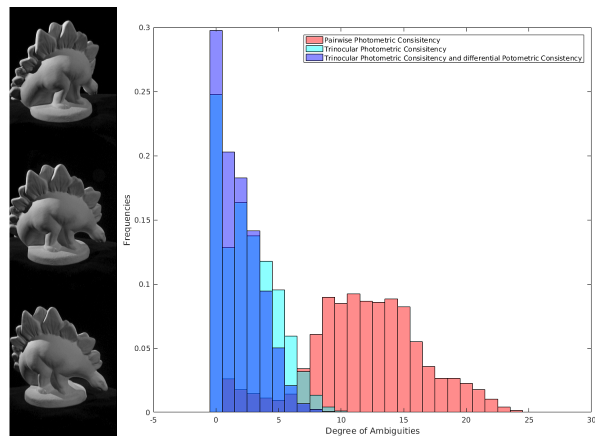
Paper ID ****

Abstract

*Photometric consistency is a key measure in establishing correspondences among views of a scene in multi-view stereo (MVS) surface reconstruction. The large correspondence ambiguity in photometric consistency led to the matching of local intensity patterns (*e.g., *using SIFT) as candidate corresponding keypoints (*e.g., Harris corners*) and this has formed a core technology for computing camera pose and for 3D reconstruction. The reconstruction in the form of an unorganized set of points and a resulting mesh is generally excellent for a texture-rich area but gaps remain in areas which lack features. Approaches for growing surface patches at 3D reconstruction points, such as PMVS, alleviate smaller gaps to the extent that surfaces can be extended. However, larger shaded areas lack features and this weakens the photometric consistency cue needed to establish dense correspondences. In this paper we introduce a novel constraint,* differential photometric consistency, *which constraints image gradients of one view from image gradients of two other views at the corresponding points. Similarly, image Hessians in one view are determined from image gradients and image Hessians in two other views. It is shown that these additional constraints reduce correspondences ambiguity and should lead to robust and more accurate reconstruction. In addition, we show how the differential geometry of a surface patch can be determined from image gradients at corresponding points from two views: (i) The surface normal is determined from image gradients in two views, and (ii) the second fundamental form of the surfaces can be determined from image Hessians in two views. This is largely a theoretical paper, but it is supported by some preliminary and illustrative experiments.*

1. Introduction

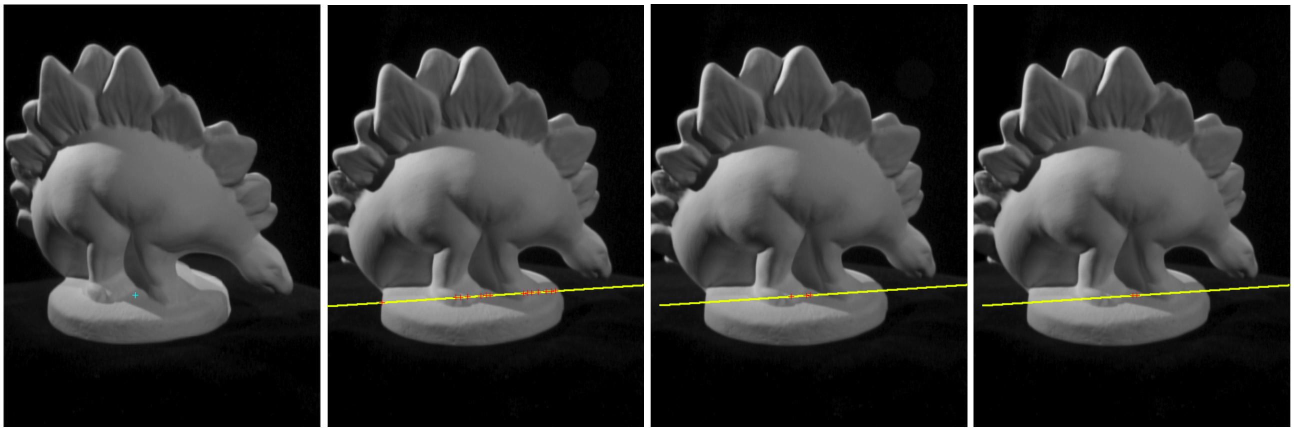
Multi-view stereo (MVS) is a key area in computer vision for the 3D reconstruction of scenes from a collection of images or from video clips with a wide range of applications such as engineering and scientific data analysis, reconstruction of urban scenes, video game industry, and film editing. The input images can come from photo collections of known sites (*e.g.,* tourist sites, urban scenes, etc.), multiple cell phone images of a scene, simultaneous multi-camera acquisition, and more recently from robots, drones or Micro Aerial Vehicles (MAVs). The accuracy of reconstruction has reached a point where MVS reconstruction can be considered as a more affordable alternative to costly laser scans. In general, the increased availability of collections of photographs from a scene and an increasing number of applications has led to significant progress in this area, but in its general setting, MVS reconstruction remains an open and challenging area.



(a) (b)

Figure 1: This figure illustrates how the use of differential photometric consistency reduces ambiguity. (a) Sample images used in experiment; (b) Three histogram represent the degree of ambiguity by showing how many ambiguous matches exist for each point under (i) photometric consistency between two images, (ii) photometric consistency among three images, and (iii) photometric and differential photometric consistency among three images.

Current MVS approaches can be generally grouped into four types, based on the type of the underlying representation of 3D structures [11]. First, in **voxel-based** ap-



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

Figure 2: The sample results of ambiguity of correspondence and corresponding epipolar line. (a) The target pixel in the first view. (b) The ambiguity in correspondence along the epipolar line using constraint (i). (c) The ambiguity in correspondence along the epipolar line using constraint (ii). (d) The ambiguity in correspondence along the epipolar line using constraint (iii). Constraints (i-iii) described in Section 4.2.

proaches, a bounding box is placed around the object and discretized, thus determining the accuracy of the reconstruction [10, 16, 12]. The location and size of the object of interested has to be roughly known in this method. Second, in the **deformable surface** approach, a surface model in the form of a polygonal mesh, or level set, is initialized, e.g. using a visual hull model, and then evolved to be optimally consistent with the data [1, 3]. Third, in the **multi-depth map** approach, the scene is initially represented by multiple depth maps for each viewpoint, and these are then fused into a single mesh representation [5, 14]. Finally, in the **local surface patch** method, object surfaces are initially represented by an unorganized collection of surface patches that are grown from corresponding features and these are then joined to form a single mesh representation [8, 6, 4]. The first two methods place restrictions on the scene as the objects need to be isolated and their initial location/size or visual hull need to be easily computable. The latter two methods are more generally applicable to more challenging scenes with clutter, partial occlusion, and self-occlusion.

A key notion underlying these methods is to recognize when a proposed reconstruction is consistent across projected views. A typical approach is to either project candidate point, patch, or volume elements onto various views and measure the consistency of expected projected intensity from the data or compare the consistency of putative correspondences between two images. The computational measure of consistency is referred to as the *photo-consistency* measure [11] which can be implemented in a number of ways. This measure is based on the Lambertian surface assumption, namely, that the observed intensities, or irradiance values, are a product of the illumination flux reaching the surface from the light source and albedo, but not a function of the viewing direction. As such, image points from different views that correspond to the same surface point should have the same intensity. Practically, a 3D point or 3D patch is consistent across views if the intensity variance across projected pixels in different views is small [7]. Similarly, an image point in view one that corresponds to an image point in view two are photometrically consistent if their intensity difference corresponds to image noise, motivating measuring sum of squared differences (SSD) or normalized cross correlation (NCC) between fixed windows around candidate corresponding points [2, 13]. More general reflection functions (BRDFs) have motivated a number of new photo-consistency metrics [17, 15, 18].

The photometric consistency generates multiple ambiguous matches. Thus it is generally a weak measure to establish dense correspondences. An alternative approach to measuring consistency of projections is to compare the patterns of intensities in two images around the candidate corresponding points. This approach relies on feature descriptors such as SIFT [9], which are typically insensitive to intensity changes, to summarize the pattern at a point. The correspondence and matching of isolated feature points using such descriptors is a core technology for computing pose between two views. The reconstruction at these isolated points results in an unorganized set of 3D points which can be joined to form a mesh, provided there is ample texture on the surface. A more recent development enables reconstruction for surfaces whose texture is not necessarily dense, by growing surface patches from reconstructed features points [4]. This has led to the very popular PMVS software. Unfortunately, the growth of a surface patch is

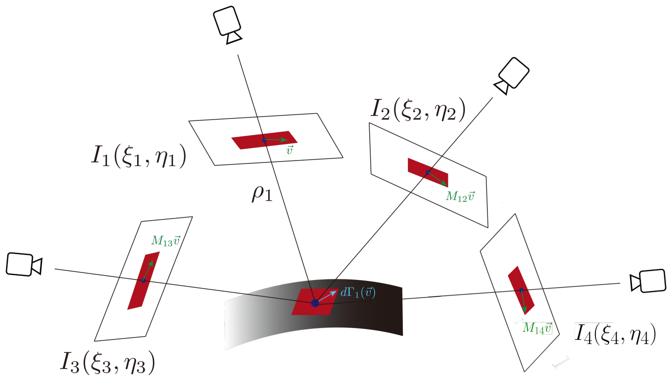


Figure 3: The multiview camera setup. Several cameras (two shown) capture images of the same shaded surface. A gradient in I_1, ∇I_1 can be predicted from a gradient ∇I_2 by mapping it to the surface and projecting it onto I_1. Conversely, the depth gradient $\nabla \rho$ can be solved for from known corresponding gradients ∇I_1 and ∇I_2.

limited and in texture-less areas PMVS is unable to generate a whole surface without leaving behind gaps. This motivates an approach to deal with the reconstruction of shaded areas.

Observe that the main difficulty in reconstructing texture-less, shaded areas is the high level of ambiguity that pairwise photometric consistency leaves behind: each point in one image simply matches too many points, Figure 2(b). The weakness of photometric consistency can be overcome to some extent by enforcing trinocular geometry: each candidate pair of corresponding points in two images defines a unique point in a third image, which is in turn expected to be photometrically consistent with both points. This significantly reduces ambiguity, as shown in Figure 2(c), but this is not sufficient as a significant level of ambiguity remains. In this paper we present a novel notion of *differential photometric consistency*, namely, that at a triplet of corresponding points across three images, the image intensity gradient at two of three points, ∇I_1 and ∇I_2 uniquely determine the third ∇I_3. Enforcing this notion of first-order "differential photometric consistency" reduces correspondence ambiguity, Figure 2(d), which in turn should lead to more robust reconstructions. We have also derived a different second-order differential photometric where the image intensity Hessian in two views H_{I_1} and H_{I_2} determine the Hessian in a third view H_{I_3}, although the use of second-order derivatives faces numerical challenges in applications.

In addition to relating differentials photometic properties in three views, we show that (i) The surface normal can be obtained from the image intensity gradient at corresponding projected points from two views, and (ii) The surface second fundamental form (which gives principal curvatures and principal directions) can be computed from the Hessian of the image at corresponding points.

A brief summary of the paper is as follows: Consider a smooth patch of surface in a scene with Lambertian reflectance and its projection to an arbitrary view. Observe that the unknown surface depth in one view is a single dimension of variability whose knowledge determines the location of the corresponding points in all other views. Thus, two corresponding points in two views define surface depth, which in turn defines all other projected points in other views. Likewise, observe that the unknown surface depth gradient, which defines the surface patch normal, is a two-dimensional unknown whose knowledge determines the image intensity gradient in all other views. Thus, we show that the image intensity gradients at two corresponding points define the surface depth gradient, or the surface normal, and as such determine all intensity gradients in all other views. This also holds for second order differentials: The Hessian of the surface depth, which defines the surface second fundamental form (which defines surface principal curvatures), is a three-dimensional unknown whose knowledge defines the Hessian of intensities in any project view. Thus, the Hessian of intensity at two corresponding points gives the surface depth second fundamental form and in turn, determines the Hessian of intensities in all other views. We can also state these in the form of first-order and second-order differential photometric constrains between triplets of corresponding points. We posit that such constraints significantly reduce ambiguity when matching points in multiview stereo and while also giving a richer form of surface geometry. We expect that these theoretical developments can be employed in a PMVS-like growth of surfaces from sparse point correspondences with the differential relations enabling the growth of large surface patches and more accurate reconstructions than current versions of PMVS. While the intent of this theoretical paper is to communicate these novel differential photometric constraints, we have done some preliminary experiments. It is our plan to implement

Table 1: Notation Table

I_i	image i
γ_i	point on image plane (in meter unit)
$\not\gamma_i$	point on image (in pixel unit)
(x_i, y_i)	the two dimensional index into I_i in meter unit
(ξ_i, η_i)	the two dimensional index into I_i in pixel unit
(X_i, Y_i, Z_i)	coordinates in the ith camera frame in $\mathbb{R}^3$
ϕ_{ij}	the map taking a point in image 1 to its corresponding point in image 2
ρ_i	the depth map $\mathbb{R}^2 \to \mathbb{R}^1$ associating an image point to its $\mathbb{R}^3$ point
Γ_i	the map $\mathbb{R}^2 \to \mathbb{R}^3$ associating an image point to its $\mathbb{R}^3$ point
K_i	the intrinsic camera matrix associated with image i
R_{ij}	a rotation matrix from the frame B_i to frame B_j
T_{ij}	a translation vector from the frame B_i to frame B_j
Λ_{ij}	The mapping from one image frame i to image frame j
Π_i	the perspective projection operator from the world frame of $\mathbb{R}^3$ onto the image plane P_i

a MVS reconstruction based on this idea in the near future.

2. First Order Analysis: From a Pair of Image Intensity Differentials to Surface Gradient

We consider N cameras viewing a scene containing a surface patch $S \in \mathbb{R}^3$, as in Figure 3. Given a point $\Gamma \in S$, define $\Gamma_i = (X_i, Y_i, Z_i)$ to be the coordinates in the coordinate frame of camera i, $i = 1, 2, ...N$. Let I_i denote the image obtained from camera i. The coordinates of two cameras are related by

$$\Gamma_j = R_{ij}\Gamma_i + T_{ij}, \tag{1}$$

where R_{ij} is a rotation matrix and T_{ij} is the translation vector from camera i to camera j. We use shorthand notation Λ_{ij} to denote this mapping: $\Lambda_{ij}(\Gamma_i) = \Gamma_j$.

The projection of the point Γ onto the image plane of camera i is denoted by $\Pi_i(\Gamma_i) = \gamma_i$, where $\gamma_i^T = (x_i, y_i, 1)$. Here, x_i and y_i are the image coordinates in meters and we assume the normalized focal length of the camera is 1. This gives

$$\Gamma_i = \rho_i \gamma_i, \tag{2}$$

where ρ_i is the depth of Γ_i in camera i. We also need to express the projected point in pixel units since our images are matrices of pixels. Let $\not\gamma_i = (\xi_i, \eta_i, 1)$ where ξ_i and η_i are the horizontal and vertical image coordinates of the projected point in pixels. These two representations of the projected point are related by the calibration matrix K_i,

$$\not\gamma_i = K_i \gamma_i, \qquad \begin{bmatrix} \xi_i \\ \eta_i \\ 1 \end{bmatrix} = K_i \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}, \tag{3}$$

$$K_i = \begin{bmatrix} F_i^x & 0 & \mu_i \\ 0 & F_i^y & \nu_i \\ 0 & 0 & 1 \end{bmatrix}, \quad K_i^{-1} = \begin{bmatrix} \frac{1}{F_i^x} & 0 & -\frac{\mu_i}{F_i^x} \\ 0 & \frac{1}{F_i^y} & -\frac{\nu_i}{F_i^y} \\ 0 & 0 & 1 \end{bmatrix} \tag{4}$$

where F_i^x and F_i^y are the pixel width and height in units of focal length, and (μ_i, ν_i) are the coordinates of the principal point in pixels, *i.e.*, the footpoint of the camera center c_i onto the image plane P_i.

The Differential Photometric Constraint: The traditional constraint for finding the correspondence between a point in one camera and a point in another is the *photometric constraint*, which states that given a Lambertian surface, the intensity at corresponding points $\not\gamma_1$ and $\not\gamma_2$ are identical:

$$I_j(\not\gamma_2) = I_i(\not\gamma_1). \tag{5}$$

Let the correspondence mapping between $\not\gamma_1$ and $\not\gamma_2$ be defined as ϕ_{12}:

$$\not\gamma_2 = \phi_{12}(\not\gamma_1). \tag{6}$$

This mapping is mediated by the surface S since $\not\gamma_1$ and $\not\gamma_2$ arise from a common surface point Γ. More specifically, $\not\gamma_1$ maps to γ_1, γ_1 maps to Γ_1, Γ_1 becomes Γ_2 under a change of basis, Γ_2 projects to γ_2, which gives $\not\gamma_2$ in pixel units:

$$\not\gamma_2 = \phi_{12}(\not\gamma_1) = K_2\left(\Pi_2\left(\Lambda_{12}\left(\Gamma_1\left(K_1^{-1}\left(\not\gamma_1\right)\right)\right)\right)\right), \tag{7}$$

where $\Gamma_i(\cdot)$ is the mapping of γ_i to Γ_1 and Π_2 is the projection of Γ_2 to γ_2.

A comment on notation: the image I_i is usually considered a mapping $\mathbb{R}^2 \to \mathbb{R}$. In this analysis, we will conveniently treat $I_i : \mathbb{R}^3 \to \mathbb{R}$ to keep all the matrices a fixed size of 3×3. The intensity at pixel location (ξ_i, η_i) is given by $I_i(\not\gamma_i)$ where $\not\gamma_i = \{\xi_i, \eta_i, 1\}$.

Our key question is how the intensity gradient $\nabla I_2^T = (\frac{\partial I_2}{\partial \xi_2}, \frac{\partial I_2}{\partial \eta_2}, 0)$ is related to the intensity gradient $\nabla I_1^T = (\frac{\partial I_1}{\partial \xi_1}, \frac{\partial I_1}{\partial \eta_1}, 0)$. If we assume that S is smooth and that Γ is not on an occluding contour, the photometric constraint will hold at every point in neighborhoods $\not\gamma_1 \in \Omega_1 \subset P_1$

and $\not\gamma_2 \in \Omega_2 \subset P_2$. We can rewrite Equation 5 as

$$I_2(\phi_{12}(\not\gamma_1)) = I_1(\not\gamma_1), \tag{8}$$

A change in $\not\gamma_1$, $d\not\gamma_1$, thus induces a change dI_1 in image 1 and dI_2 in image 2 that are related via the chain rule:

$$dI_2 \circ d\phi_{12}(d\not\gamma_1) = dI_1(d\not\gamma_1). \tag{9}$$

Equivalently:

$$\nabla I_2 \, d\phi_{12} = \nabla I_1. \tag{10}$$

Thus, to compare intensity gradients across images, we must solve for $d\phi_{ij}$, where ϕ_{ij} is defined in Equation 7.

Proposition 2.1. *The image gradients at tow corresponding points $\not\gamma_i$ and $\not\gamma_j$ from calibrated cameras i and j can be related if the depth ρ_i and ρ_j and normalized depth gradient $\frac{\nabla\rho_i}{rho_i}$ are given, specially,*

$$\nabla I_1 = d\phi_1 2 \nabla I_2. \tag{11}$$

The expression for $d\phi_{ij}$ is $dK_j \circ d\Pi_j \circ d\Lambda_{ij} \circ d\Gamma_i \circ dK_i^{-1}$. It can be explicitly reduced to

$$d\phi_{ij} = \frac{\rho_i}{\rho_j}\begin{bmatrix} F_j^x & 0 & -(\xi_j - \mu_j) \\ 0 & F_j^y & -(\eta_j - \nu_j) \\ 0 & 0 & 0 \end{bmatrix} R_{ij} \begin{bmatrix} \frac{1}{F_i^x} & 0 & 0 \\ 0 & \frac{1}{F_i^y} & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{12}$$

$$+ \frac{\rho_i}{\rho_j}\begin{bmatrix} F_j^x & 0 & -(\xi_j - \mu_j) \\ 0 & F_j^y & -(\eta_j - \nu_j) \\ 0 & 0 & 0 \end{bmatrix} R_{ij} \begin{bmatrix} \frac{\xi_i - \mu_i}{F_i^x} \\ \frac{\eta_i - \nu_i}{F_i^y} \\ 1 \end{bmatrix}$$

$$\frac{1}{\rho_i}\begin{bmatrix} \frac{\partial\rho_i}{\partial\xi_i} & \frac{\partial\rho_i}{\partial\eta_i} & 0 \end{bmatrix}.$$

Proof. We differentiate Equation 7, yielding

$$d\phi_{12} = dK_j \circ d\Pi_j \circ d\Lambda_{ij} \circ d\Gamma_i \circ dK_i^{-1} \tag{13}$$

Each differential mapping can be represented by a Jacobian matrix which we calculate below.

The differential of K_i and K_i^{-1}: Using Equation 4, the mapping is K_i defined as

$$K_i\left(\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}\right) = \begin{bmatrix} F_i^x x_i + \mu_i \\ F_i^y y_i + \nu_i \\ 1 \end{bmatrix} \tag{14}$$

Thus, its Jacobian and inverse Jacobian are:

$$dK_i = \begin{bmatrix} F_i^x & 0 & 0 \\ 0 & F_i^y & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad dK_i^{-1} = \begin{bmatrix} \frac{1}{F_i^x} & 0 & 0 \\ 0 & \frac{1}{F_i^y} & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{15}$$

The differential of Γ_i: Using Equation 2, $\Gamma_i(\cdot)$ is the mapping of γ_i to Γ_1:

$$\Gamma_i\left(\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}\right) = \begin{bmatrix} \rho_i x_i \\ \rho_i y_i \\ \rho_i \end{bmatrix}. \tag{16}$$

Its Jacobian is:

$$\begin{aligned} d\Gamma_i &= \begin{bmatrix} \frac{\partial\rho_i}{\partial x_i}x_i + \rho_i & \frac{\partial\rho_i}{\partial y_i}x_i & 0 \\ \frac{\partial\rho_i}{\partial x_i}y_i & \frac{\partial\rho_i}{\partial y_i}y_i + \rho_i & 0 \\ \frac{\partial\rho_i}{\partial x_i} & \frac{\partial\rho_i}{\partial y_i} & 0 \end{bmatrix} \\ &= \rho_i\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}\begin{bmatrix} \frac{\partial\rho_i}{\partial x_i} & \frac{\partial\rho_i}{\partial y_i} & 0 \end{bmatrix} \\ &= \rho_i\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} + K_i^{-1}\begin{bmatrix} \xi_i \\ \eta_i \\ 1 \end{bmatrix}\begin{bmatrix} \frac{\partial\rho_i}{\partial\xi_i} & \frac{\partial\rho_i}{\partial\eta_i} & 0 \end{bmatrix} dK_i. \end{aligned} \tag{17}$$

The differential of Λ_{ij}: Using Equation 1, the mapping Λ_{ij} is defined as

$$\Lambda_{ij}\left(\begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix}\right) = R_{ij}\begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} + T_{ij}. \tag{18}$$

Its Jacobian is

$$d\Lambda_{ij} = R_{ij}. \tag{19}$$

The differential of Π_i: Π_i is the projection of Γ_i to γ_i:

$$\Pi_i\left(\begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix}\right) = \begin{bmatrix} \frac{X_i}{Z_i} \\ \frac{Y_i}{Z_i} \\ 1 \end{bmatrix}. \tag{20}$$

Its Jacobian is simplified using $\begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} = \rho_i\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}$ and $\not\gamma_i = K_i^{-1}\gamma_i$:

$$d\Pi_i = \begin{bmatrix} \frac{1}{Z_i} & 0 & -\frac{X_i}{Z_i^2} \\ 0 & \frac{1}{Z_i} & -\frac{Y_i}{Z_i^2} \\ 0 & 0 & 0 \end{bmatrix} = \frac{1}{\rho_i}\begin{bmatrix} 1 & 0 & -\frac{\xi_i - \mu_i}{F_i} \\ 0 & 1 & -\frac{\eta_i - \nu_i}{F_i} \\ 0 & 0 & 0 \end{bmatrix}. \tag{21}$$

The differential of $d\phi_{ij}$: Combining the previous expressions, we get:

$$d\phi_{ij} = \frac{\rho_i}{\rho_j}\begin{bmatrix} F_j^x & 0 & -(\xi_j - \mu_j) \\ 0 & F_j^y & -(\eta_j - \nu_j) \\ 0 & 0 & 0 \end{bmatrix} R_{ij} \begin{bmatrix} \frac{1}{F_i^x} & 0 & 0 \\ 0 & \frac{1}{F_i^y} & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{22}$$

$$+ \frac{\rho_i}{\rho_j}\begin{bmatrix} F_j^x & 0 & -(\xi_j - \mu_j) \\ 0 & F_j^y & -(\eta_j - \nu_j) \\ 0 & 0 & 0 \end{bmatrix} R_{ij} \begin{bmatrix} \frac{\xi_i - \mu_i}{F_i^x} \\ \frac{\eta_i - \nu_i}{F_i^y} \\ 1 \end{bmatrix}$$

$$\frac{1}{\rho_i}\begin{bmatrix} \frac{\partial\rho_i}{\partial\xi_i} & \frac{\partial\rho_i}{\partial\eta_i} & 0 \end{bmatrix}$$

We have thus derived the constraint of intensity gradients among two images, shown in Equation 13. The only unknown involved in the constraint is the depth gradient:

$$\nabla \rho_i^T = (\frac{\partial \rho_i}{\partial \xi_i}, \frac{\partial \rho_i}{\partial \eta_i}) \tag{23}$$

With two images, one can only solve for this depth gradient. A third image allows us to verify whether that depth gradient is plausible. Thus, as described below, one needs a triplet of intensity gradients to decide if the first order image neighborhoods of three corresponding points are consistent.

Three Images Model: Consider three calibrated images, I_i, I_j and I_k containing three corresponding points, denoted by γ_i, γ_j and γ_k.

Proposition 2.2. *The image gradients at corresponding points γ_i, γ_j and γ_k from three calibrated cameras i, j, k need to satisfy the following constraint.*

$$\nabla I_i = \nabla I_k \Bigg(\frac{\rho_i}{\rho_k} \begin{bmatrix} F_k^x & 0 & -(\xi_k - \mu_k) \\ 0 & F_k^y & -(\eta_k - \nu_k) \\ 0 & 0 & 0 \end{bmatrix} R_{i,k} \begin{bmatrix} \frac{1}{F_i^x} & 0 & 0 \\ 0 & \frac{1}{F_i^y} & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{24}$$

$$+ \frac{\rho_i}{\rho_k} \begin{bmatrix} F_k^x & 0 & -(\xi_k - \mu_k) \\ 0 & F_k^y & -(\eta_k - \nu_k) \\ 0 & 0 & 0 \end{bmatrix} R_{i,k} \begin{bmatrix} \frac{\xi_i - \mu_i}{F_i^x} \\ \frac{\eta_i - \nu_i}{F_i^y} \\ 1 \end{bmatrix}$$

$$\frac{1}{\rho_i} \begin{bmatrix} \frac{\partial \rho_i}{\partial \xi_i} & \frac{\partial \rho_i}{\partial \eta_i} & 0 \end{bmatrix} \Bigg),$$

where

$$\frac{1}{\rho_i} \begin{bmatrix} \frac{\partial \rho_i}{\partial \xi_i} & \frac{\partial \rho_i}{\partial \eta_i} & 0 \end{bmatrix} = \tag{25}$$

$$\frac{\frac{\rho_j}{\rho_i} \nabla I_i - \nabla I_j \begin{bmatrix} F_j^x & 0 & -(\xi_j - \mu_j) \\ 0 & F_j^y & -(\eta_j - \nu_j) \\ 0 & 0 & 0 \end{bmatrix} R_{i,j} \begin{bmatrix} \frac{1}{F_i^x} & 0 & 0 \\ 0 & \frac{1}{F_i^y} & 0 \\ 0 & 0 & 0 \end{bmatrix}}{\nabla I_j \begin{bmatrix} F_j^x & 0 & -(\xi_j - \mu_j) \\ 0 & F_j^y & -(\eta_j - \nu_j) \\ 0 & 0 & 0 \end{bmatrix} R_{i,j} \begin{bmatrix} \frac{\xi_i - \mu_i}{F_i^x} \\ \frac{\eta_i - \nu_i}{F_i^y} \\ 1 \end{bmatrix}}$$

Proof. Using Equation 10 on each pair of images, we obtain a system of two equations:

$$\begin{cases} \nabla I_2 \, d\phi_{12} = \nabla I_1 \\ \nabla I_3 \, d\phi_{13} = \nabla I_1 \end{cases} \tag{26}$$

Solve for $\frac{1}{\rho_i} \begin{bmatrix} \frac{\partial \rho_i}{\partial \xi_i} & \frac{\partial \rho_i}{\partial \eta_i} & 0 \end{bmatrix}$ from the first of these equations, getting Equation 25. The second of these equations can then be expressed as Equation 24 with this substitution. ∎

From this corollary, we see that the intensity gradient in the third image is predicted from the intensity gradients, calibration parameters, and depth ratios $\frac{\rho_i}{\rho_j}, \frac{\rho_i}{\rho_k}$ in the other two images. We now compute those depth ratios.

Calculation of depth ratio $\frac{\rho_i}{\rho_j}$: Starting with the 3D point transformation:

$$\rho_j \gamma_j = \rho_i R_{ij} \gamma_i + T_{ij} \tag{27}$$

Take the cross product with T_{mn} on both sides,

$$\rho_j (\gamma_j \times T_{ij}) = \rho_i (R_{ij} \gamma_i \times T_{ij}) \tag{28}$$

If the corresponding points exactly satisfy epipolar constraints, then the two vectors $(\gamma_j \times T_{ij})$ and $(R_{ij}\gamma_i \times T_{ij})$ are proportional. (In practice, these values will be close to proportional.) We solve for the depth ratio

$$\frac{\rho_i}{\rho_j} (R_{ij} \gamma_i \times T_{ij}) = (\gamma_j \times T_{ij}) \tag{29}$$

So the ratio can be solved by minimize the following optimization problem:

$$\min_{\frac{\rho_i}{\rho_j}} || \frac{\rho_i}{\rho_j} (R_{ij} \gamma_i \times T_{ij}) - (\gamma_j \times T_{ij}) || \tag{30}$$

The solution of this problem is

$$r_{ij} = \frac{\rho_i}{\rho_j} = \frac{a_1 b_1 + a_2 b_2 + a_3 b_3}{a_1^2 + a_2^2 + a_3^2} \tag{31}$$

where $(R_{ij}\gamma_i \times T_{ij}) = [a_1, a_2, a_3]^T$ and $(\gamma_j \times T_{ij}) = [b_1, b_2, b_3]^T$.

3. Second Order Analysis

We now repeat the above analysis in the second order case. We consider local image information up to second order (Hessians) and show a pair of image neighborhoods in two separate images implies the depth Hessian.

Theorem 3.1. *Let $I_1(\xi_1, \eta_1)$ and $I_2(\xi_2, \eta_2)$ be two images of the same smooth surface S with depth function ρ. For any pair of corresponding points $\gamma_1 \in I_1, \gamma_2 \in I_2$, there is a map $\bar{\Psi}$ that calculates the Hessian of ρ from the image gradients $\nabla I_1, \nabla I_2$ and image Hessians H_{I_1}, H_{I_2}:*

$$\bar{\Psi}_{1,2} : \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^{2\times2} \times \mathbb{R}^{2\times2} \to \mathbb{R}^{2\times2} \tag{32}$$

$$\bar{\Psi}_{1,2}(\nabla I_1, \nabla I_2, H_{I_1}, H_{I_2}) = H_\rho \tag{33}$$

Similarly, we will have the corresponding corollary as before.

Corollary 3.1. *One can also predict the Hessian in the third image from the 2nd order expansion in the first two images. There exists a $\bar{\chi}_{1,2} : \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^{2\times2} \times \mathbb{R}^{2\times2} \to \mathbb{R}^{2\times2}$*

$$\bar{\chi}_{1,2}(\nabla I_1, \nabla I_2, H_{I_1}, H_{I_2}) = H_{I_3} \tag{34}$$

Sketch of proof of Theorem 3.1. The full proof is included in the supplementary material. Here, we just show plausibility. We take Equation **??** and regarding it as a linear equation on the vector fields (∇I_1 and ∇I_2), we can differentiate it again with respect to ξ_1. (η_1 is analogous.) Note that the basis of each matrix does not change with respect to the differentiation variable, so we can apply a product rule for matrices. We show the derivation for the first row of the Hessian. Let $D_{\xi_1}(\cdot)$ represent the differentiation operator w.r.t ξ_1:

$$\begin{pmatrix} I_{1,\xi_1\xi_1} \\ I_{1,\xi_1\eta_1} \end{pmatrix} = D_{\xi_1}\left(\nabla I_1\right) \tag{35}$$

$$\begin{aligned} &= D_{\xi_1}\left(dK_2 \cdot d\Pi_2 \cdot dR \cdot d\Gamma_1 \cdot dK_1^{-1}\right)\nabla I_2 \\ &+ M_{12} D_{\xi_1}\left(\nabla I_2\right) \end{aligned} \tag{36}$$

We now analyze each of the two terms $D_{\xi_1}(dK_2 \cdot d\Pi_2 \cdot dR \cdot d\Gamma_1 \cdot dK_1^{-1})$ and $D_{\xi_1}(\nabla I_2)$.

$$D_{\xi_1}(dK_2 \cdot d\Pi_2 \cdot dR \cdot d\Gamma_1 \cdot dK_1^{-1}) \tag{37}$$

$$\begin{aligned} &= dK_2 \cdot d\Pi_2 \cdot dR \cdot D_{\xi_1}(d\Gamma_1) \cdot dK_1^{-1} \\ &+ dK_2 \cdot D_{\xi_1}(d\Pi_2) \cdot dR \cdot d\Gamma_1 \cdot dK_1^{-1} \end{aligned} \tag{38}$$

The other terms resulting from the product rule are 0, since the other matrices are constant. Note that $D_{\xi_1}(d\Gamma_1)$ is a linear function of the second order depth derivatives $\{\rho_{\xi\xi}, \rho_{\xi\eta}\}$ of the surface. Also, note that $D_{\xi_1}(d\Pi_2)$ is nonzero but known from the 3D world position of point s. Thus, the terms 38 can be separated into known quantities and a known matrix multiplied by the second derivatives $\begin{pmatrix}\rho_{\xi\xi} & \rho_{\xi\eta}\end{pmatrix}$.

To calculate $D_{\xi_1}(\nabla I_2)$, we can split it into two components of ∇I_2. Consider the first component $D_{\xi_1}(I_{2,\xi_2})$ This is simply a directional derivative of a scalar quantity, but we have to be careful with the direction! Writing $\vec{w_1} = M_{12}\begin{bmatrix}1 \\ 0\end{bmatrix}$, which can be calculated from the first order analysis, then:

$$D_{\xi_1}(I_{2,\xi_2}) = \vec{w} \cdot \begin{bmatrix} I_{2,\xi_2\xi_2} \\ I_{2,\xi_2\eta_2} \end{bmatrix} \tag{39}$$

Thus,

$$D_{\xi_1}(\nabla I_2) = \vec{w_1}^T \begin{bmatrix} I_{2,\xi_2\xi_2} & I_{2,\xi_2\eta_2} \\ I_{2,\xi_2\eta_2} & I_{2,\eta_2\eta_2} \end{bmatrix} \tag{40}$$

$$= \vec{w_1}^T H_{I_2} \tag{41}$$

where H_{I_2} is the second image Hessian. Substituting Equations 38 and 41 into 36, we find two equations consisting of known quantities and a linear function of $\begin{bmatrix}\rho_{\xi\xi} & \rho_{\xi\eta}\end{bmatrix}$. Thus, the variables $\rho_{\xi\xi}$ and $\rho_{\xi\eta}$ are solvable from the image Hessians.

The proof of Corollary 3.1 is very similar to Corollary 2.2. Once the second order depth variables $\{\rho_{\xi\xi}, \rho_{\xi\eta}, \rho_{\eta\eta}\}$ are known, the third image Hessian can be computed from pulling the first image Hessian up to the third order surface neighborhood and then projecting it down into P_3.

4. Experiments and Illustrative Examples

A Theoretical Experiment To experimentally verify the analysis and especially Equation **??**, we constructed symbolic depth maps imaging the same surface from different views. See Figure 4. We set $\rho_1(\xi_1, \eta_1) = \xi_1 + \eta_1$ on the domain $\xi_1, \eta_1 > 0$ and then chose a second view direction 45 degrees offset and solved for $\rho_2(\xi_2, \eta_2)$ by inverting Equation **??**. (We chose this simple function for ρ_1 so that this inversion can be done symbolically in Mathematica.) These two depth maps are shown in the first row of Figure 4. Note that ρ_2 is not linear. We then lit the surface with a single light source and found symbolic expressions for the two Lambertian images corresponding to the two views. This images are shown in the second row of Figure 4.

By having symbolic expressions for both the depth maps and shaded images, we were able to compute every differential matrix in Section 2 precisely at any point. (Compare to Figure **??** where the matrices were prone to error due to noisy image derivatives.) In particular, we checked that Equation **??** held up to machine error precision. **A Practical Experiment** In this set of experiments,

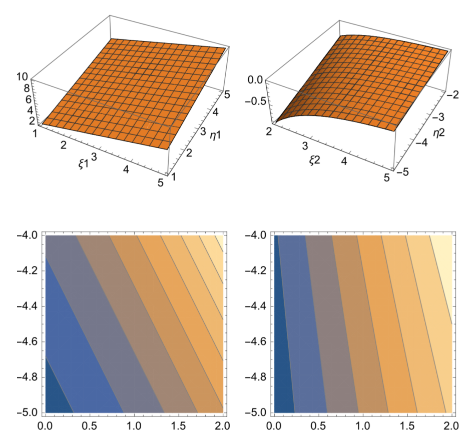


Figure 4: Top row, left: A planar depth map ρ_1. Top row, right: A parametrically defined depth map ρ_2 corresponding to the same surface and a shifted camera pose. Bottom row: A local image 1 patch constructed from imaging the surface described by ρ_1 (left) mapped exactly to a local image 2 patch (right) under a computed M_{12}.

the effectiveness of differential photometric consistency in reducing the ambiguity of the correspondence was examined. Specifically, three views of the same scene are selected from a calibrated dataset, Figure 1(a). We used an image from the Middlebury dataset [11]. Three different constraints are used to generate the correspondences: (i) **Pairwise photometric consistency**: for each point in the first view, all points lying on the corresponding epipolar line in the second image whose intensity difference was within a threshold τ_1 were found. (We used $\tau_1 = 15$.) The size of the set composed of such points represents the degree of ambiguity in finding the true correspondence. This is done for all points in the first views and a histogram of frequency vs. size of ambiguity is shown as the green histogram in Figure 1(b); (ii) **Trinocular photometric consistency**: for each point in the first image, and for each candidate match in the second image, a corresponding point in a third image is computed by intersection of the corresponding epipolar lines. In addition, for each triplet of points, the pairwise photometric constraint is enforced by requiring the intensity difference between any of the three pairs of points to be within the threshold τ_1. The size of the remaining set represents the degree of ambiguity as shown in the cyan histogram of Figure 1(b); (iii) **Trinocular photometric consistency and differential photometric consistency**: For each triplet satisfying the condition of constraints represented in (ii), we also enforced a differential photometric constraint, represented in Equation **??**, among the three gradients; namely, the predicted gradient in the first image (using gradients in the second and thrid images) and actual gradient should lie within a threshold τ_2. (We used $\tau_2 = 50$.) The size of the remaining set is the degree of ambiguity as shown in the blue histogram of Figure 1(b). Observe that (a) the histogram using constraint (iii) has less ambiguity than (ii) which is in turn better than (i), and (b) the average ambiguity drops by using constraint (iii) vs. (ii) which in turn is better than (i).

This process can also be illustrated on an exemplar point shown in Figure 2(a) as a cyan cross. The set of all matches satisfying constraints (i), (ii) and (iii) are shown as red crosses in Figure 2(b), (c) and (d), respectively. Observe how the numerous matches satisfying constraint (i), averaging 11.98 per pixel, are reduced in number as constraint (ii) is applied, averaging 2.55 per pixel. The number of matches is reduced even further by enforcing constraint (iii), resulting in 1.82 matches per pixel.

5. Conclusion

In this paper, we introduced *differential photometric consistency* so the image gradient in one view can be uniquely determined by two other views in the context of multiview stereo. Preliminary experiments verified the correctness of the theory and showed the effectiveness in reducing the degree of ambiguity of correspondence. In addition, we also described *second-order differential photometric consistency* which denotes the analogous map defining the image Hessian in the third view from a pair of Hessians in other views. These additional photometric constraints reduce the ambiguity associated with multiview stereo, which is especially relevant when viewing smoothly shaded surfaces.

References

[1] C. H. Esteban and F. Schmitt. Silhouette and stereo fusion for 3d object modeling. *Computer Vision and Image Understanding*, 96(3):367–392, 2004.

[2] O. Faugeras and R. Keriven. Variational principles, surface evolution, pdes, level set methods, and the stereo problem. *IEEE Transactions on Image Processing*, 7(3):336–344, 1998.

[3] Y. Furukawa and J. Ponce. Carved visual hulls for image-based modeling. *International Journal of Computer Vision*, 81(1):53–67, 2009.

[4] Y. Furukawa and J. Ponce. Accurate, dense, and robust multiview stereopsis. *IEEE transactions on pattern analysis and machine intelligence*, 32(8):1362–1376, 2010.

[5] M. Goesele, B. Curless, and S. M. Seitz. Multi-view stereo revisited. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 2402–2409. IEEE, 2006.

[6] M. Habbecke and L. Kobbelt. Iterative multi-view plane fitting. In *Int. Fall Workshop of Vision, Modeling, and Visualization*, 2006.

[7] K. N. Kutulakos and S. M. Seitz. A theory of shape by space carving. *International Journal of Computer Vision*, 38(3):199–218, 2000.

[8] M. Lhuillier and L. Quan. A quasi-dense approach to surface reconstruction from uncalibrated images. *IEEE transactions on pattern analysis and machine intelligence*, 27(3):418–433, 2005.

[9] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[10] J.-P. Pons, R. Keriven, and O. Faugeras. Multi-view stereo reconstruction and scene flow estimation with a global image-based matching score. *International Journal of Computer Vision*, 72(2):179–193, 2007.

[11] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Computer vision and pattern recognition, 2006 IEEE Computer Society Conference on*, volume 1, pages 519–528. IEEE, 2006.

[12] S. N. Sinha, P. Mordohai, and M. Pollefeys. Multi-view stereo via graph cuts on the dual of an adaptive tetrahedral mesh. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE, 2007.

[13] S. Soatto, A. Yezzi, and H. Jin. Tales of shape and radiance in multiview stereo. In *Proceedings Ninth IEEE International Conference on Computer Vision*.

[14] C. Strecha, R. Fransens, and L. Van Gool. Combined depth and outlier estimation in multi-view stereo. In *Computer Vi-*

sion and Pattern Recognition, 2006 IEEE Computer Society Conference on, volume 2, pages 2394–2401. IEEE, 2006.

[15] A. Treuille, A. Hertzmann, and S. M. Seitz. Example-based stereo with general brdfs. In *European Conference on Computer Vision*, pages 457–469. Springer, 2004.

[16] G. Vogiatzis, P. H. Torr, and R. Cipolla. Multi-view stereo via volumetric graph-cuts. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 391–398. IEEE, 2005.

[17] R. Yang, M. Pollefeys, and G. Welch. Dealing with textureless regions and specular highlights-a progressive space carving scheme using a novel photo-consistency measure. In *Proceedings of the Ninth IEEE International Conference on Computer Vision-Volume 2*, page 576. IEEE Computer Society, 2003.

[18] T. Yu, N. Xu, and N. Ahuja. Shape and view independent reflectance map from multiple views. *International journal of computer vision*, 73(2):123–138, 2007.