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 constrains 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.

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-
approaches, a bounding box is placed around the object and discretized, thus determining the accuracy of the reconstruction. The location and size of the object of interest 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. Finally, 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. Third, 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.

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 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. 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. More general reflection functions (BRDFs) have motivated a number of new photo-consistency metrics.

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, 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. This has led to the very popular PMVS software.
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, \( \nabla I_1 \) and \( \nabla I_2 \) uniquely determine the third \( \nabla I_3 \). 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 photometric 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 dimensional variable 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 multi-view 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
2. First Order Analysis: From a Pair of Image Differences 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 $i$, $i = 1, 2, \ldots, N$. Let $I_i$ denote the image obtained from camera $i$. The coordinates of two cameras are related by

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

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

The projection of the point $\Gamma$ onto the image plane of camera $i$ is denoted by $\Pi_i(\Gamma_i) = \gamma_i$, where $\gamma_i = (x_i, y_i, 1)$. 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

$$\Gamma_i = \rho_i \gamma_i,$$

where $\rho_i$ is the depth of $\Gamma_i$ in camera $i$. We also need to express the projected point in pixel units since our images are matrices of pixels. Let $\gamma_i' = (\xi_i, \eta_i, 1)$ where $\xi_i$ and $\eta_i$ 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 $K_i$.

$$\gamma_i' = K_i \gamma_i,$$

$$K_i = \begin{bmatrix} F_x & 0 & \mu_i \\ 0 & F_y & \nu_i \\ 0 & 0 & 1 \end{bmatrix}, \quad K_i^{-1} = \begin{bmatrix} \frac{1}{F_x} & 0 & -\frac{\mu_i}{F_x} \\ 0 & \frac{1}{F_y} & -\frac{\nu_i}{F_y} \\ 0 & 0 & 1 \end{bmatrix}. \tag{4}$$

where $F_x$ and $F_y$ are the pixel width and height in units of focal length, and $(\mu_i, \nu_i)$ 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 $\gamma_1$ and $\gamma_2$ are identical:

$$I_j(\gamma_2) = I_i(\gamma_1). \tag{5}$$

Let the correspondence mapping between $\gamma_1$ and $\gamma_2$ be defined as $\phi_{12}$:

$$\gamma_2 = \phi_{12}(\gamma_1). \tag{6}$$

This mapping is mediated by the surface $S$ since $\gamma_1$ and $\gamma_2$ arise from a common surface point $\Gamma$. More specifically, $\gamma_1$ maps to $\gamma_1$, $\gamma_1$ maps to $\Gamma_1$, $\Gamma_1$ becomes $\Gamma_2$ under a change of basis, $\Gamma_2$ projects to $\gamma_2$, which gives $\gamma_2$ in pixel units:

$$\gamma_2 = \phi_{12}(\gamma_1) = K_2 (\Pi_2 (\Lambda_{12} (\Gamma_1 (K_1^{-1}(\gamma_1))))), \tag{7}$$

where $\Gamma_i(\cdot)$ is the mapping of $\gamma_i$ to $\Gamma_i$ and $\Pi_2$ is the projection of $\Gamma_2$ to $\gamma_2$.

A comment on notation: the image $I_i$ is usually considered a mapping $\mathbb{R}^2 \rightarrow \mathbb{R}$. In this analysis, we will conveniently treat $I_i : \mathbb{R}^3 \rightarrow \mathbb{R}$ to keep all the matrices a fixed size of $3 \times 3$. The intensity at pixel location $(\xi_i, \eta_i)$ is given by $I_i(\gamma_i)$ where $\gamma_i = (\xi_i, \eta_i, 1)$.

Our key question is how the intensity gradient $\nabla I_2^T = \left(\frac{\partial I_2}{\partial x_2}, \frac{\partial I_2}{\partial y_2}, 0\right)$ is related to the intensity gradient $\nabla I_1^T = \left(\frac{\partial I_1}{\partial x_1}, \frac{\partial I_1}{\partial y_1}, 0\right)$. If we assume that $S$ is smooth and that $\Gamma$ is not on an occluding contour, the photometric constraint will hold at every point in neighborhoods $\gamma_1 \in \Omega_1 \subset P_1$ and $\gamma_2 \in \Omega_2 \subset P_2$.

Table 1: Notation Table

| $I_i$ | image i |
| $\gamma_i$ | point on image plane (in meter unit) |
| $\gamma_i'$ | point on image (in pixel unit) |
| $(x_i, y_i)$ | the two dimensional index into $I_i$, in meter unit |
| $(\xi_i, \eta_i)$ | the two dimensional index into $I_i$, in pixel unit |
| $(X_i, Y_i, Z_i)$ | coordinates in the ith camera frame in $\mathbb{R}^3$ |
| $\phi_{ij}$ | the map taking a point in image 1 to its corresponding point in image 2 |
| $\rho_i$ | the depth map $\mathbb{R}^2 \rightarrow \mathbb{R}^1$ associating an image point to its $\mathbb{R}^3$ point |
| $\Gamma_i$ | the map $\mathbb{R}^2 \rightarrow \mathbb{R}^3$ 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 $B_j$ |
| $\Lambda_{ij}$ | The mapping from one image frame i to image frame j |
| $\Pi_i$ | the perspective projection operator from the world frame of $\mathbb{R}^3$ onto the image plane $P_i$ |
and $\gamma_2 \in \Omega_2 \subset P_2$. We can rewrite Equation 5 as

$$ I_2(\phi_{12}(\gamma_1)) = I_1(\gamma_1), \quad (8) $$

A change in $\gamma_1$, $d\gamma_1$, thus induces a change $dI_1$ in image 1 and $dI_2$ in image 2 that are related via the chain rule:

$$ dI_2 \circ d\phi_{12}(d\gamma_1) = dI_1(d\gamma_1). \quad (9) $$

Equivalently:

$$ \nabla I_2 \, d\phi_{12} = \nabla I_1. \quad (10) $$

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

**Proposition 2.1.** The image gradients at two corresponding points $\gamma_i$ and $\gamma_j$ from calibrated cameras $i$ and $j$ can be related if the depth $\rho_i$ and $\rho_j$ and normalized depth gradient $\nabla \rho_i / \nabla \rho_j$ are given, specially,

$$ \nabla I_1 = d\phi_{ij} \nabla I_2. \quad (11) $$

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

$$ d\phi_{ij} = \frac{\rho_i}{\rho_j} \left[ \begin{array}{cc} F_j & 0 \\ 0 & F_j \end{array} \right] \left[ \begin{array}{c} -\xi_j - \mu_j \\ -\eta_j - \nu_j \end{array} \right] \, R_{ij} \left[ \begin{array}{c} 1 \\ 0 \\ 0 \end{array} \right] \left[ \begin{array}{c} 0 \\ \frac{\xi_i - \mu_i}{F_i} \\ \frac{\eta_i - \nu_i}{F_i} \end{array} \right] \quad (12) $$

$$ \quad + \left( \begin{array}{c} 0 \\ 0 \\ 0 \end{array} \right) $$

$$ \quad + \frac{1}{\rho_i} \left[ \begin{array}{c} \frac{\partial \rho_i}{\partial \xi_i} \\ \frac{\partial \rho_i}{\partial \eta_i} \end{array} \right] \left[ \begin{array}{c} 0 \\ 0 \end{array} \right]. $$

Proof. We differentiate Equation 7 yielding

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

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

$$ K_i \left( \begin{array}{c} x_i \\ y_i \end{array} \right) = \left[ \begin{array}{c} F_i x_i + \mu_i \\ F_i y_i + \nu_i \end{array} \right] \quad (14) $$

Thus, its Jacobian and inverse Jacobian are:

$$ dK_i = \left[ \begin{array}{cc} F_i & 0 \\ 0 & F_i \end{array} \right], \quad dK_i^{-1} = \left[ \begin{array}{cc} 1 & 0 \\ 0 & \frac{1}{F_i} \end{array} \right] \quad (15) $$

**The differential of $\Gamma_i$:** Using Equation 2, $\Gamma_i(\cdot)$ is the mapping of $\gamma_i$ to $\Gamma_1$:

$$ \Gamma_i \left( \begin{array}{c} x_i \\ y_i \end{array} \right) = \left[ \begin{array}{c} \rho_i x_i \\ \rho_i y_i \end{array} \right]. \quad (16) $$

Its Jacobian is:

$$ d\Gamma_i = \left[ \begin{array}{cc} \frac{\partial \rho_i}{\partial x_i} x_i + \rho_i & \frac{\partial \rho_i}{\partial y_i} x_i \\ \frac{\partial \rho_i}{\partial x_i} y_i + \rho_i & \frac{\partial \rho_i}{\partial y_i} y_i \\ 0 & 0 \end{array} \right] \quad (17) $$

$$ = \rho_i \left[ \begin{array}{c} 1 & 0 & 0 \\ 0 & 1 & 0 \end{array} \right] + \left[ \begin{array}{c} x_i \\ y_i \end{array} \right] \left[ \begin{array}{c} \frac{\partial \rho_i}{\partial x_i} \frac{\partial \rho_i}{\partial y_i} \end{array} \right] \quad (18) $$

Thus, its Jacobian is:

$$ d\Lambda_{ij} = R_{ij} \quad (19) $$

**The differential of $\Pi_i$:** $\Pi_i$ is the projection of $\Gamma_i$ to $\gamma_i$:

$$ \Pi_i \left( \begin{array}{c} X_i \\ Y_i \end{array} \right) = \left[ \begin{array}{c} x_i \\ z_i \\ 1 \end{array} \right]. \quad (20) $$

Its Jacobian is simplified using $[X_i, Y_i, Z_i] = \rho_i [x_i, y_i, 1]$ and $\gamma_i = K_i^{-1} \gamma_i$:

$$ d\Pi_i = \left[ \begin{array}{cccc} 0 & 0 & 0 \\ 0 & 0 & -\frac{x_i}{z_i} \\ 1 & 0 & 0 \end{array} \right] = \frac{1}{\rho_i} \left[ \begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{array} \right] \quad (21) $$

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

$$ d\phi_{ij} = \frac{\rho_i}{\rho_j} \left[ \begin{array}{cc} F_j & 0 \\ 0 & F_j \end{array} \right] \left[ \begin{array}{c} -\xi_j - \mu_j \\ -\eta_j - \nu_j \end{array} \right] \, R_{ij} \left[ \begin{array}{c} 1 \\ 0 \\ 0 \end{array} \right] \left[ \begin{array}{c} 0 \\ \frac{\xi_i - \mu_i}{F_i} \\ \frac{\eta_i - \nu_i}{F_i} \end{array} \right] \quad (22) $$

$$ + \frac{1}{\rho_i} \left[ \begin{array}{c} \frac{\partial \rho_i}{\partial \xi_i} \\ \frac{\partial \rho_i}{\partial \eta_i} \end{array} \right] \left[ \begin{array}{c} 0 \\ 0 \end{array} \right] $$
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:

$$\nabla \rho_i^T = \left( \frac{\partial \rho_i}{\partial x_i}, \frac{\partial \rho_i}{\partial y_i} \right)$$

(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, $I_1, I_2$ and $I_3$ containing three corresponding points, denoted by $\gamma_i, \gamma_j$ and $\gamma_k$.

**Proposition 2.2.** The image gradients at corresponding points $\gamma_i, \gamma_j$ and $\gamma_k$ from three calibrated cameras $i, j, k$ need to satisfy the following constraint.

$$\nabla I_i = \nabla I_j \left( \frac{\rho_j}{\rho_k} \begin{bmatrix} F_{x} & 0 & -\xi & \mu_k \\ 0 & F_{y} & -\nu_k & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \right)$$

(24)

$$= \frac{\rho_i}{\rho_j} \left( R_{i,j} \right) \frac{\nabla I_i}{\rho_k}$$

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

$$\Psi_{1,2} : \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}^{2 \times 2}$$

(32)

$$\Psi_{1,2}(\nabla I_1, \nabla I_2, H_{I_1}, H_{I_2}) = H_\rho$$

(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 $\tilde{\chi}_{1,2} : \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}^{2 \times 2}$

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

(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 3.1 and regarding it as a linear equation on the vector fields (∇I₁ and ∇I₂), we can differentiate it again with respect to ξ₁. (η₁ 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 Dξ₁(·) represent the differentiation operator w.r.t ξ₁:

\[
\begin{align*}
(1_{1}, ξ₁, η₁, \eta_{1}, m) &= D_{ξ₁}(∇I₁) \\
&= D_{ξ₁}(dK₂ \cdot dΠ₂ \cdot dR \cdot dΓ₁ \cdot dK^{-1}_{1}) \nabla I₂ \\
&= M₁₂D_{ξ₁}(∇I₂)
\end{align*}
\]

We now analyze each of the two terms Dξ₁(dK₂ \cdot dΠ₂ \cdot dR \cdot dΓ₁ \cdot dK⁻¹₁) and Dξ₁(∇I₂).

\[
D_{ξ₁}(dK₂ \cdot dΠ₂ \cdot dR \cdot dΓ₁ \cdot dK⁻¹₁) = dK₂ \cdot dΠ₂ \cdot dR \cdot D_{ξ₁}(dΓ₁) \cdot dK⁻¹₁ \\
+ dK₂ \cdot D_{ξ₁}(dΠ₂) \cdot dR \cdot dΓ₁ \cdot dK⁻¹₁ (37)
\]

The other terms resulting from the product rule are 0, since the other matrices are constant. Note that Dξ₁(∂Γ₁) is a linear function of the second order depth derivatives \{ρξξ, ρξη\} of the surface. Also, note that Dξ₁(dΠ₂) is nonzero but known from the 3D world position of point s. Thus, the terms can be separated into known quantities and a known matrix multiplied by the second derivatives \{ρξξ, ρξη\}.

To calculate Dξ₁(∇I₂), we can split it into two components of ∇I₂. Consider the first component Dξ₁(I₂,ξ₂). This is simply a directional derivative of a scalar quantity, but we have to be careful with the direction! Writing \(\vec{w}_1 = M₁₂ \begin{bmatrix} 1 \\ 0 \end{bmatrix}\), which can be calculated from the first order analysis, then:

\[
D_{ξ₁}(I₂,ξ₂) = \vec{w}_1 \cdot \begin{bmatrix} I₂,ξ₂,ξ₂ \\ I₂,ξ₂,η₂ \end{bmatrix}
\]

Thus, the effectiveness of differential photometric consistency in reducing the ambiguity of the correspondence was examined. Specifically, three views of the same scene are

![Figure 4: Top row, left: A planar depth map ρ₁. Top row, right: A parametrically defined depth map ρ₂ corresponding to the same surface and a shifted camera pose. Bottom row: A local image 1 patch extracted from an image of the surface described by ρ₁ (left) mapped exactly to a local image 2 patch (right) under a computed M₁₂.](image-url)
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 $\tau_1$ 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 $\tau_1$. 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 third images) and actual gradient should lie within a threshold $\tau_2$. (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


