PHENOM: Interest Points on Photometric Normal Maps

Reza Sabzevari, Eren Alak and Davide Scaramuzza †

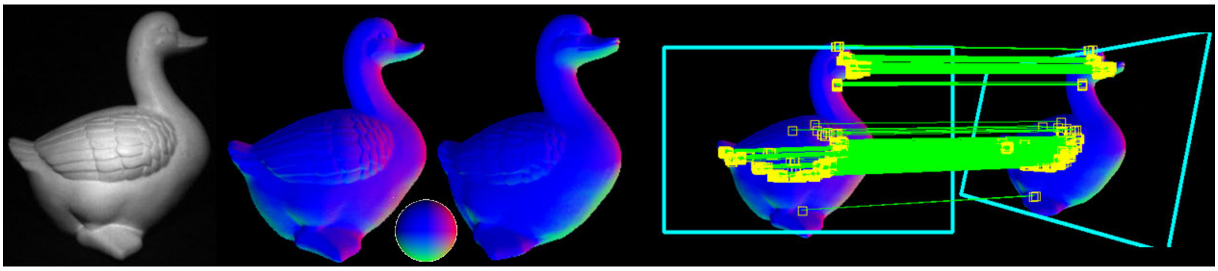

Figure 1: *PHENOM Features: An image from the Duck sequence (left), Normal maps for two different views (middle), and Phenom features with matching between two views (right).*

Abstract

This paper introduces a novel method for extracting features and matching points on images of texture-less surfaces. Feature points are extracted from surface normal maps recovered by Photometric Stereo. Such sparse matching will help to register high detail 3D surfaces reconstructed from multiple view images. Moreover, the geometric constraints imposed by multiple views can be utilized to correct the geometric ambiguity in photometric reconstruction. Experiments show the performance of proposed interest points in matching texture-less objects. Comparison against texture-based interest points shows that the proposed features based on normal maps perform effectively.

Categories and Subject Descriptors (according to ACM CCS): I.4.8 [Image Processing and Computer Vision]: Scene Analysis—Photometry; I.4.5 [Image Processing and Computer Vision]: Reconstruction—Transform methods

1. Introduction

One of the main challenges in 3D perception with monocular cameras is the limited performance of available methods imposed by the scene properties. Geometric approaches need observable motion through the aperture which is usually obtained by moving the camera. In such cases the scene needs to be highly textured to allow matching the points between images and estimating the depth of these points. Apparently, for those parts of the scene that are poor in texture correspondences between images are not possible to be accurately extracted. Therefore, the depth for points belonging to such parts of the scene can not be estimated [PFS]. Differently, with photometric approaches—which need fixed-viewpoint images with different lighting conditions—the surface points' orientations can be estimated accurately independent from the surface intensity (i.e. texture), but the depth estimation is subject to geometric ambiguities [SDBM12]. Exploiting both photometric and geometric properties of the scene has been attracted interest of researchers in recent years [JK, SDBM]. Extracting geometric information without reliable detection of interest points is not possible. This is more challenging with texture-less scenes where stable interest points are not easy to extract and match. This work introduces a novel definition for interest points based on vector map of surface normals. In the following sections detection and matching of such interest points are described and their performance is presented and discussed.

† Robotics and Perception Group, University of Zurich, Switzerland (http://rpg.ifi.uzh.ch/)

2. PHENOM Features

The most properties for interest points are accuracy, distinctiveness and repeatability. The normal map obtained by Photometric Stereo consists of homogeneous regions—where neighbouring normals are similar—and heterogeneous regions—where neighbouring normals significantly differ from each other.

DOI: 10.2312/egp.20151040

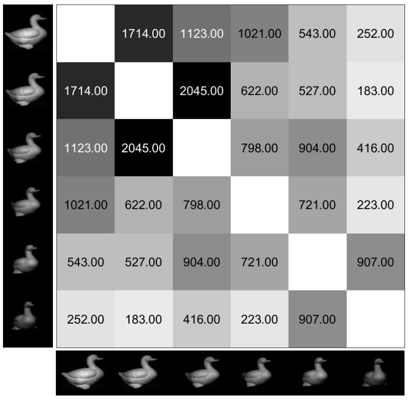

Figure 2: *Number of inlier obtained by matching different view points. Outlier are removed by geometric verification.*

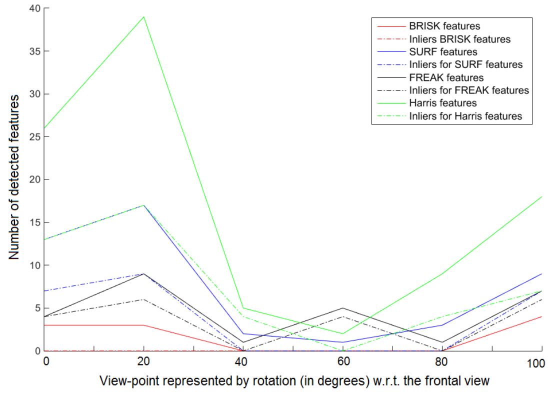

Figure 3: *Number of detected features using texture-based feature on multiple views and slight change of lighting.*

Phenom features are defined as a set of distinctive predefined patches, called *template patches*. Such patches consist of matrices of normal unit vectors that correspond to distinctive surface patches, i.e. corner, edge, sink and source of the normal field. the Phenom features have three main properties:

- Since the surface normal vectors are defined in the object's reference frame, the corresponding vectors from normal maps of two different views do not have the same direction, the Phenom features should be rotation invariant. This feature is warranted by aligning central normal vector of each patch to the patch normal vector before matching them.
- In order to handle scale changes caused by the camera movement, the detection of Phenom features is done in multiple levels of scales.
- Rotation invariance about the Z-axis—defined by the camera—should be taken into account. For this reason the descriptor holds the sorted orientation of normal vectors.

To find template features on the normal map, every patch is transformed in a way that the normal vector in the middle is aligned with the camera's Z-axis and then compared with the set of template features by rotating the patch about the Z-axis in 45 degrees steps and in 3 different scale levels. To detect Phenom features, a similarity measurement is defined as

$$similarity\ \ measure = \sum_{i=1}^{N} \sum_{j=1}^{N} \vec{n}_{rot,ij} \cdot \vec{n}_{f,ij} \qquad (1)$$

to measure the distance between the current patch on the normal map and each template patch. In (1), N is the patch size, $\vec{n}_{rot,ij}$ is the aligned normal vector at position (i, j) on the normal map and $\vec{n}_{f,ij}$ is the normal vector on the template patch. Such a metric is also used in matching Phenom features between multiple views.

3. Experiments

To evaluate the performance of Phenom features, we have compared against well-known texture-based point features, i.e. BRISK, SURF, Harris and FREAK. One may argue that if the object is texture-less then there is no point to compare with texture-based point features. For this reason, we have chosen several test cases with some observable distinctive parts on the surface to be able to use texture-based point features. Indeed, such distinctive parts on the surface are apparent because of the cast shadows posed by some parts of the surface. These surface parts look distinctive to human eye but they are subject to convex-concave illusion. This means that, the convexity or concavity of such surface parts are not distinguishable and by changing the light direction a convex area may turn to concave or vice versa. Fig. 2 shows the number of matches in different views after removing the outliers by geometric verification, i.e. RANSAC. Fig. 3 shows the number of detected interest points and inliers using texture-based features.

4. Conclusions

This paper presents a novel interest point defined on normal maps motivated by matching texture-less objects in multiple views. The presented preliminary results show the impact of such feature points defined on the surface normals. This method can be used for registering multiple dense surfaces obtained by Photometric Stereo and also provides cues to correct the geometric ambiguities in photometric reconstructions. Further improvements for this method can be considered by substituting the predefined set of template features with a smart way to extract interest surface patches. This can be done by using interest point detectors on normal field based on normal filed decomposition methods.

References

[JK] JOSHI N., KRIEGMAN D.: Shape from varying illumination and viewpoint. In *ICCV'07*. 1

[PFS] PIZZOLI M., FORSTER C., SCARAMUZZA D.: Remode: Probabilistic, monocular dense reconstruction in real time. In *ICRA'14*. 1

[SDBM] SABZEVARI R., DEL BUE A., MURINO V.: Multi-view photometric stereo using semi-isometric mappings. In *3DIMPVT'12*. 1

[SDBM12] SABZEVARI R., DEL BUE A., MURINO V.: Combining structure from motion and photometric stereo: A piecewise formulation. Eurographics. 1