Novel View Synthesis using Needle-Map Correspondence

Philip L. Worthington
Department of Computation, UMIST, Manchester, UK
plw@co.umist.ac.uk

Abstract

Interest in view interpolation and novel view synthesis is growing. In this paper we show how dense correspondence can be found between needle-maps generated using shape-from-shading, which in turn can be used to generate new needle-maps. From these we can produce novel intermediate views, and also estimates of how each intermediate view would look under different lighting conditions. The approach offers the prospect of creating large sets of realistic views of a scene under different viewing and lighting conditions from a small number of original images.

1 Introduction

Interest in view interpolation and novel view synthesis has grown rapidly over the last decade. Much of the motivation for this stems from the needs of virtual- and tele- reality, and computer games. Storing a small number of 2-D views of a real object, and estimating how it would look from new viewpoints, has clear attractions compared to creating explicit, accurate 3-D models. Moreover, even a good 3-D model usually requires texture mapping in order to produce realistic-looking images. In contrast, interpolating views of real objects allows realistic images to be created without a model.

View-based object recognition researchers have also explored recognising new views by interpolation or alignment. Although there are various realizations of the idea, the general principal is to compute a compact representation of the 2D appearance of 3D objects under multiple viewing and illumination conditions. For each object, a potentially huge number of images must be be collected for different combinations of pose and lighting. However, it is possible to reduce the number of views needed by using representations that are stable to changes in viewpoint and illumination. Alternatively, new views can be estimated from a small set of sample views, and subsequently matched. For instance, Ullman and Basri have shown how multiple views may be combined for recognising intermediate object poses [16]. This is taken to the limit by Vetter, Beymer and Poggio [7, 2], who use a single face image to generate many novel views under variable pose, although they also use prior knowledge of the class of face images in the generation of virtual views via warping.

In this paper, we take two of the classic techniques from computer vision - correspondence matching and SFS - and demonstrate that by interpolating pairs of needle-maps it is possible to generate realistic intermediate images. Moreover, since our approach involves producing an intermediate 2.5D representation, it is also possible to re-illuminate both the original and interpolated images over a small but useful range of illuminant directions.

7 18

The research reported here is strongly influenced by the work of Scharstein [12, 10], which describes the use of dense correspondence for view interpolation. Scharstein's basic correspondence algorithm, which uses gradient similarity to find matches, is adapted to work using the surface normal estimates supplied by SFS. We subsequently demonstrate that it is possible to find good, dense correspondences on the basis of surface normal estimates. These can be used to generate intermediate needle-maps which can then be re-illuminated under the same conditions as the original images, or with different lighting.

We demonstrate the approach for stereo image pairs where the calibration information is unknown or unused. In doing so, we assume the goal is to provide realistic, plausible intermediate images that are convincing to a human observer, rather than quantitatively accurate new views.

2   Background

In most computer graphics and VR applications, objects must be represented as 3D models, often with texture mapping used to improve realism. Constructing such models from passive sensors, particularly cameras, is potentially very attractive. However, the results of image-based model construction, with the possible exception of stereo systems, have typically proved unspectacular. This is particularly true of shape-from-shading (SFS), which has received a great deal of attention since Horn first formulated the problem in the vision literature [5]. Whilst it is clear that one of the motivations of early SFS research was to enable a 3-D representation to be derived from a single image, the difficulties encountered in achieving accurate and robust surface height recovery have proved a serious obstacle to progress in this direction.

In the face of the difficulties in recovering reliable 3D information, many researchers have abandoned the goal of constructing an explicit 3D model using passive methods, and have instead concentrated on direct view synthesis, e.g. [1, 13, 19, 12]. Several recent papers have reported the use of shading information in a view synthesis context. For example, Zhao and Chellappa [22] used SFS to synthesise frontal views of human faces from other views, essentially correcting for the effects of pose and illumination in a face recognition context. Meanwhile, the issue of re-illuminating an image under novel lighting conditions has attracted attention from Belhumeur, Kriegman and Georghiades [3, 4] and Levoy and Hanrahan [6].

We have shown in [19] that, using a robust and reliable SFS scheme [21], we can perform crude interpolation between two needle-maps. There, our motivation was to expand a set of images to produce novel views for eigenspace-based recognition, with the aim of producing a manifold for an object from a limited set of views. The view synthesis process in [19] did not make use of correspondence information.

The novel contribution of this paper is to develop a true view synthesis approach using SFS, by making use of dense correspondence between needle-maps. We adapt the correspondence matching approach of Scharstein [10] to work directly with the needle-maps. Subsequently, we generate a new needle-map by transporting normals to an interpolated position. This can then be illuminated, either with the same lighting as the original images. Similarly, the needle-maps of the original images can also be re-illuminated, thus allowing a large set of images of to be synthesised from the original image pair.

3 Needle-Map Correspondence

The surface normals used for view synthesis in this paper are generated by our curvature-consistency SFS algorithm, introduced in [21], although any SFS scheme with reasonable performance could potentially be used. The algorithm performs well on a wide variety of real-world images, avoiding oversmoothing and proving stable to viewpoint changes. Although use of the image irradiance equation as a hard constraint makes the recovered normals susceptible to image noise in terms of the angle subtended between the normal and light-source direction, the direction of the normals in the image plane is relatively stable. Smoothing the image also helps mitigate noise effects.

Ragheb and Hancock [9] have recently reported additional development of this approach to include a specular model. Such enhancements could improve the view synthesis process reported here, since most real images deviate from Lambert's law.

Much of the literature on stereo correspondence focuses on feature-based methods which yield only a sparse correspondence map. Although it is possible to produce a dense mapping by interpolation, view synthesis is generally considered to require a dense map from the outset. Pixel-based matching is possible, for example using the Sum of Squared Differences (SSD) algorithm, but this is notoriously sensitive to local noise and global lighting variations. Robust methods for obtaining dense correspondence have been proposed recently, starting with Scott and Longuet-Higgins [14]. Pilu [8] enhances this algorithm by penalising matches that are not in corresponding regions of each image. Birchfield and Tomasi have applied dynamic programming techniques and a modified cost function that accomodates partial occlusion [15], while Scharstein [12, 10] proposes an enhanced version of the SSD algorithm, using gradient matching to remove the dependence on absolute brightness values.

Matching on the basis of surface normal estimates introduces new considerations, some of which are dependent on the general camera and lighting arrangement. The direction of the true normal associated with a given surface point is fixed relative to the object, not the viewer. To compensate for the effect of this on the recovered needle-map, we would need to know or estimate the rotation, the latter being a non-trivial task. In light of difficulties encountered in obtaining wide baseline stereo pairs, and to avoid the need to model rotation, in this paper we opt to ignore such issues and concentrate on typical stereo pairs for which any rotation is small. However, another approach would be to use relative rather than absolute normal directions, or to find correspondences using the image gradient or correlation methods, and only introduce the surface normal estimates during the view-synthesis process. However, by matching on the basis of surface normals, we demonstrate the stability of the needle-maps, and mitigate against the interpolation of significantly differing normals.

We also opt to ignore the complex issue of camera calibration, and concentrate on the uncalibrated matching of pixels or needles. Although we fully expect that a reliable calibration step would greatly improve our correspondence estimates, we would be forced to rely on a feature-based correspondence step in the calibration process in order to retain generality of application. Instead, we take advantage of recent advances in uncalibrated stereo matching [14, 12, 10, 8] and adopt a "brute-force" approach to finding correspondences.

3.1 Finding Dense Correspondences

The Scharstein [12, 10] approach matches image locations on the basis of gradient directions, making it a natural candidate for extension to needle-maps. The algorithm aims to accumulate evidence for each pixel in image I_a with each pixel in I_b within the allowable disparity range, in order to yield a disparity map $\mathbf{d}_{ab}$. The process is repeated to find $\mathbf{d}_{ba}$. A smoothing step is usually applied to the evidence values for a given disparity, in order to improve local consistency; i.e. neighbouring pixels are encouraged to have similar disparities. Gaussian smoothing is currently used, with the same kernel size applied for both gradient and normal matching, but there is clear scope for improving the results by applying the apparatus of robust statistics, or relaxation techniques, to reduce the problem of boundary blurring, in which pixels from a different object in a scene are assigned the same disparity as neighbouring pixels belonging to a different object [12].

3.2 Evidence Measure

The evidence measure defined by Scharstein [12, 10] is based on the local image gradients, $\mathbf{g}$, and aims to capture both the similarity of the gradients and the confidence in the match using:

$$e_{ab}(i,j) = \tag{1}$$
$$\max_{\Delta x, \Delta y} \left\{ - |\mathbf{g}_a(i,j) - \mathbf{g}_b(i+\Delta x, j+\Delta y)| + \frac{|\mathbf{g}_a(i,j)| + |\mathbf{g}_b(i+\Delta x, j+\Delta y)|}{2} \right\}$$

The first term of the evidence measure describes the similarity between the two gradient vectors under consideration, while the second term gives a confidence based on the mean length of the two vectors. Thus, weak gradients result in low confidence regardless of their similarity.

When matching needle-maps, the normals are unit vectors so the lengths are not relevant. Thus, we can change the evidence measure to concentrate on the similarity between the normals. To do this, we adopt the following revised evidence measure:

$$e_{ab}(i,j) = \max_{\Delta x, \Delta y} \frac{e^{\mathbf{N}_a(i,j).\mathbf{N}_b(i+\Delta x, j+\Delta y)} - e^{0.5}}{e^1 - e^{0.5}} \tag{2}$$

where $\mathbf{N}_a$ and $\mathbf{N}_b$ are a pair of needle-maps. The use of the exponential aims to encourage good matches and penalise poor matches. We scale the measure to give a maximum value of $+1$, although this formulation produces a minimum value of slightly less than -1. Figure 1 illustrates the properties of this evidence measure.

4 View Synthesis

Our correspondence process based on $\mathbf{N}_1$ and $\mathbf{N}_2$ will recover a pair of disparity maps $\mathbf{d}_{12}$ and $\mathbf{d}_{21}$. Each element of $\mathbf{d}_{12}$ describes the displacement, in pixels, from the normal in $\mathbf{N}_1$ to the corresponding normal in $\mathbf{N}_2$, and vice-versa. Clearly, the two disparity maps should be roughly symmetrical if our correspondence algorithm is successful. If the
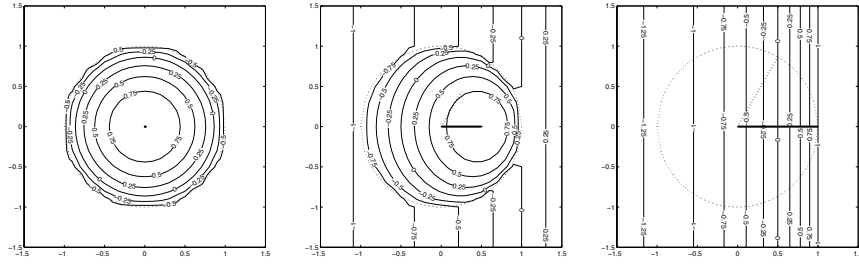
Figure 1: Evidence measure used for surface normal matching. The thick solid line illustrates the projection of a surface normal in $\mathbf{N}_a$ onto the image plane (in the left-hand case, the normal is perpendicular to the image plane). The contours illustrate the variations in the evidence measure for normals in $\mathbf{N}_b$ with different projections into the image plane. The dotted line in each case highlights the evidence value associated with a normal in $\mathbf{N}_b$ subtending an identical angle to the vertical as the normal in $\mathbf{N}_a$, but with projection onto the image plane at $60°$ to it.

intermediate needle-map is denoted $\mathbf{N}_3$, we can adapt Wolberg's [17] forward mapping:

$$\mathbf{N}_1(i,j) \rightarrow \mathbf{N}_{3,1}\left(i + X_s \mathbf{d}_{12_x}(i,j), j + Y_s \mathbf{d}_{12_y}(i,j)\right) \tag{3}$$

$$\mathbf{N}_2(i,j) \rightarrow \mathbf{N}_{3,2}\left(i + (1 - X_s)\mathbf{d}_{21_x}(i,j), j + (1 - Y_s)\mathbf{d}_{21_y}(i,j)\right) \tag{4}$$

where X_s and Y_s define an intermediate viewpoint between the original views. Typically, we concentrate on the mid-view, where $X_s = Y_s = 0.5$.

The two intermediate needle-maps need to be combined. For the purpose of this paper we simply use the following approach:

$$\mathbf{N}_3(i,j) = \begin{pmatrix} \cos\phi\sin\theta \\ \sin\phi\sin\theta \\ \cos\theta \end{pmatrix} \quad \text{where} \quad \phi = \frac{\phi_1 + \phi_2}{2}, \quad \theta = \frac{\theta_1 + \theta_2}{2}$$

$$\text{and} \quad \phi_1 = \arctan\frac{\mathbf{N}_{1_y}(i,j)}{\mathbf{N}_{1_x}(i,j)}, \quad \theta_1 = \arccos\mathbf{N}_{1_z}(i,j) \tag{5}$$

and similarly for ϕ_2 and θ_2.

4.1   Process Summary

- Apply SFS to recover needle-maps from both images
- Find dense correspondence between needle-maps and find the disparities
- Map needles to intermediate positions using the disparities to produce a pair of intermediate needle-maps
- Take the mean of the pair of intermediate needle-maps, on a normal-by-normal basis, to create a new needle-map
- Reilluminate the new needle-map to produce a novel view

5   Experiments

In the following experiments we aim to demonstrate that we can generate realistic interpolated views using needle-maps and re-illumination.

5.1 Correspondence and View Synthesis

We illustrate the correspondence results for a stereo pair of images of a Mars rock in Figure 2. It is clear that the quality of the correspondences is similar between the gradient and normals matching, although without ground-truth it is not possible to make a quantitative performance assessment. Note that in both approaches, some of the matches are between vectors which differ significantly. These matches could be eliminated by thresholding on the evidence measure, but would lead to more sampling holes in the interpolated views.
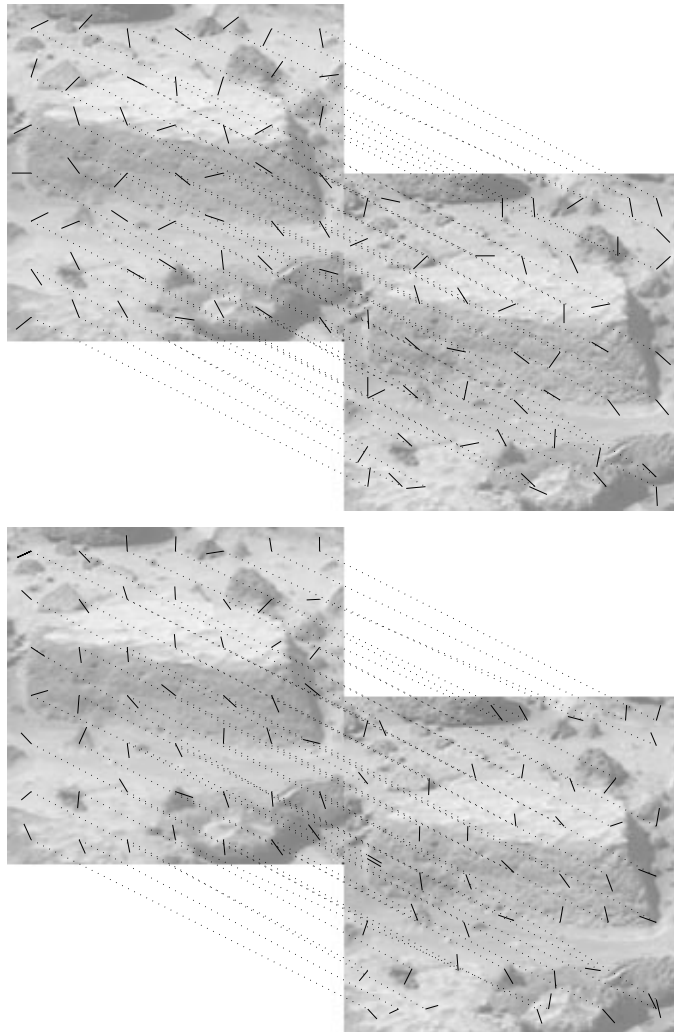

Figure 2: Top: matches between pairs of normalised gradient vectors. Bottom: matches between pairs of surface normals. Only a small fraction of matches are shown for clarity.

In Figure 3 we illustrate the results of interpolating between a pair of stereo images of the surface of Mars. We see that both the gradient- and needle-map-based interpola-

tion processes produce realistic intermediate views. Figure 5 illustrates the needle-map interpolations for three other stereo image pairs. Note that the needle-map correspondence process works even though the images are not ones which would normally have SFS applied to them, due to the good stability of the SFS process used.

It is clear that the interpolated images include hole artefacts. These are due to a mixture of sampling gaps, and the presence of scene points that are visible in one view but not the other. Small sampling gaps can be addressed by interpolating neighbouring normals, whereas holes due to occlusion and discontinuities need more careful treatment. The two cases can be distinguished by considering that a large difference in the disparity estimate from each image indicating a partial-occlusion hole [12]. In this case, interpolation will cause discontinuities to be smoothed, so extrapolation from the view in which the region is visible would be more appropriate.
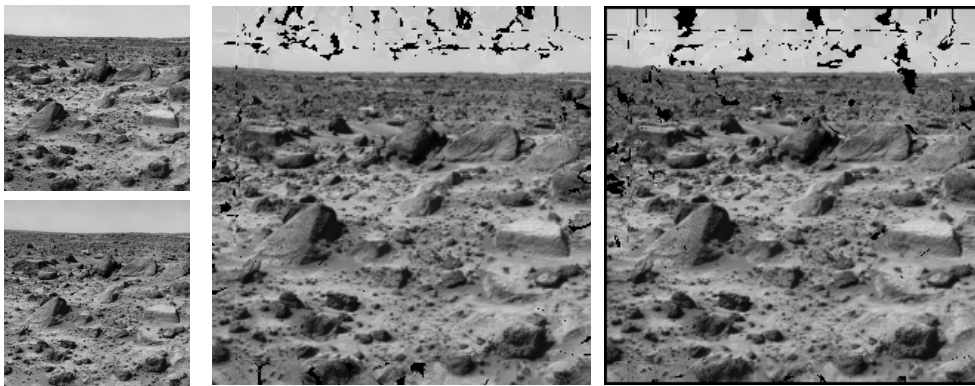

Figure 3: Left column: a stereo image pair of Mars. Interpolated images using gradient (middle) and needle-map (right) correspondence.

The use of SFS introduces the potential for noise sensitivity in the recovery of surface normals, which may adversely affect the correspondence matching. In practice, our SFS scheme [21] is sufficiently stable that good normals can be obtained by applying light Gaussian smoothing prior to matching. In Figure 4 we illustrate this by synthesising views from corrupted versions of the Mars stereo pair.

5.2   Reillumination

As the previous section illustrates, it is possible to generate realistic novel views of a scene using needle-maps. This does not in itself represent a significant advance on direct image interpolation methods, and scarcely justifies the overhead introduced by SFS. However, if we subsequently re-illuminate the interpolated needle-map with different lighting conditions, we can generate images which are not obtainable using image interpolation. To achieve this, we again assume Lambertian reflectance and constant albedo, so that the image intensity is simply given by the dot product of the normal estimate and the new light-source direction. More complex lighting and reflectance models should yield more realistic results.

Figure 6 shows the results of re-illuminating the intermediate needle-map from the Mars stereo pair under different lighting conditions. Although use of the Lambertian
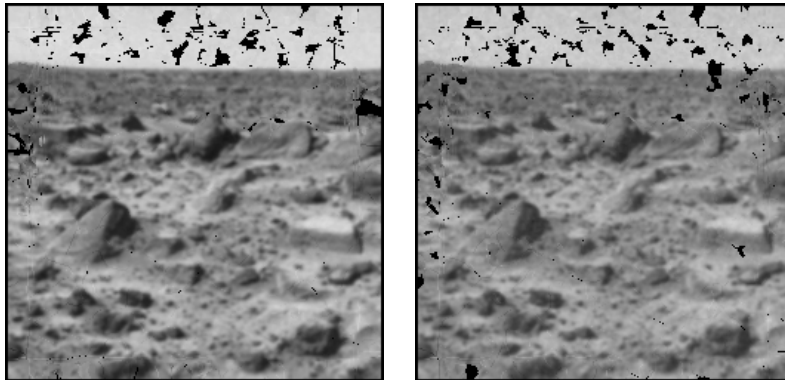
Figure 4: Top row: interpolated views using normal matching after addition of Gaussian noise with (left) $\sigma_n^2 = 10$ and (right) $\sigma_n^2 = 20$ grey levels. Gaussian smoothing with $\sigma = 1$ pixel is applied before SFS recovery to reduce noise sensitivity.
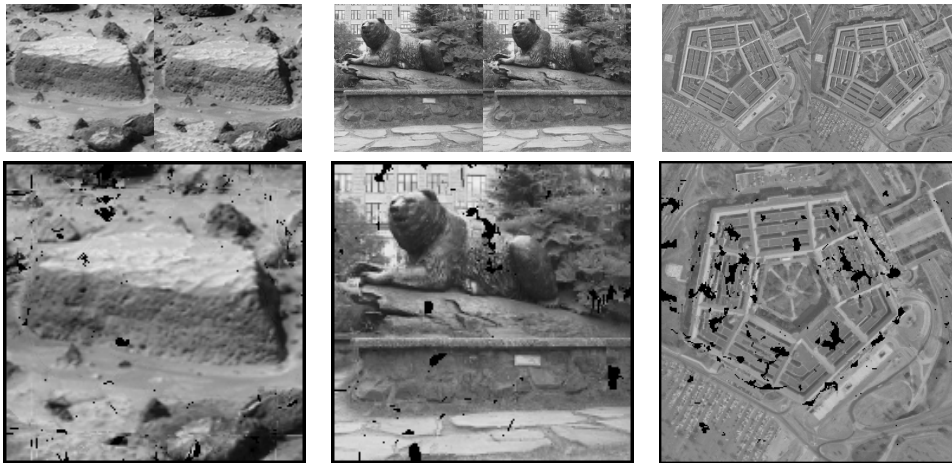

Figure 5: Top row: some stereo pair images. Bottom row: interpolated images generated using needle-map interpolation. Bear images courtesy of John R. Thurston.

model causes the process to break down for large changes in lighting direction, it is nonetheless clear that we can obtain realistic novel views under different lighting conditions which would convince a human observer if presented in a moving sequence.

6   Conclusions and Future Work

We have taken initial steps towards demonstrating that correspondence between a pair of needle-maps can be established for use in view synthesis. Even using a very simple interpolation process, and ignoring many of the complications that arise in multiple-view representations of scenes, realistic views are produced. Moreover, interpolating needle-maps enables generation a much larger set of realistic images from a given pair of original

Figure 6: Re-illumination of detail from the interpolated Mars needle-map by, from left to right, a light source at $-10°$, $-5°$, $0°$, $+5°$ and $+10°$ to the original light source along the x direction.

images.

The scope for direct enhancements to the early methods presented here is wide. Most urgent is the need to add a robust calibration step to the process, in order to rectify the images and needle-maps and allow epipolar constraints to be used. The change in normal directions due to viewpoint changes must also be modelled and predicted. Currently, we assume that these are sufficiently small that they can be neglected, but this is clearly not the case for wider baseline stereo, and for non-stereo applications such as interpolating between views of a rotating object. Once this incorporated, it will be possible to provide quantitative assessment of the interpolation quality using widely-adopted datasets featuring multiple views of objects. There is also scope to consider higher-order evidence measures, such as relative normal direction or local curvature.

To generate realistic novel views without artefacts, we need to add hole-filling methods to the process. Since the needle-maps describe local surface orientation of each view, it is anticipated that where partial occlusion is identified, the ability to extrapolate from a single view to fill the resulting holes will be greatly enhanced compared to raw image interpolation.

References

[1] Avidan, S. and Sashua, A. (1997) Novel View Synthesis in Tensor Space, *Proc. IEEE Conf. on Computer Vision and Pattern Recognition,* pp. 1034-1040.

[2] Beymer, D. and Poggio, T. (1995) Face Recognition from One Example View, A.I. Memo 1536, A.I. Laboratory, Massachusetts Institute of Technology.

[3] Georghiades, A.S., Belhumeur, P.N. and Kriegman, D.J. (1999) Illumination-based Image Synthesis: Creating Novel Images under Differing Pose and Lighting, *Proc. IEEE Workshop on Multi-view Modeling and Analysis of Visual Scenes.*

[4] Georghiades, A.S., Belhumeur, P.N. and Kriegman, D.J. (2000) From Few to Many: Generative Models for Recognition under Variable Pose and Illumination, *Proc. IEEE Int. Conf. on Automatic Face and Gesture Recognition.*

[5] Horn, B.K.P. (1975) Obtaining Shape from Shading Information. In Winston, P.H. (ed.), *The Psychology of Computer Vision,* McGraw Hill, NY, pp.115-155.

[6] Levoy, M. and Hanrahan, P. (1996) Light Field Rendering, *Computer Graphics (SIGGRAPH),* pp. 31-42.

[7] Poggio, T. and Vetter, T. (1992) Recognition and Structure from one 2D Model View: Observations on Prototypes, Object Classes and Symmetries, Memo No. 1347, A.I. Lab, MIT.

[8] Pilu, M. (1997) Uncalibrated Stereo Correspondence by Singular Value Decomposition, *Proc. Computer Vision and Pattern Recognition.*

[9] Ragheb H. and Hancock E. R. (2001) Separating Lambertian and Specular Reflectance Components using Iterated Conditional Modes, *Proc. British Machine Vision Conference.*

[10] Scharstein, D. (1994) Matching Images by Comparing their Gradient Fields, *Proc. Int. Conf. Pattern Recognition,* Vol. 1, pp. 572-575.

[11] Scharstein, D. and Szeliski, R. (1996) Stereo Matching with Non-linear Diffusion, *Proc. Computer Vision and Pattern Recognition,* pp. 343-350.

[12] Scharstein, D. (1997) *View Synthesis using Stereo Vision,* Ph.D Thesis, Cornell University.

[13] Seitz, S.M. and Dyer, C.R. (1996) View Morphing, *Computer Graphics (SIGGRAPH'96),* pp. 21-30.

[14] Scott, G. and Longuet-Higgins, H. (1991) An Algorithm for Associating the Features of Two Patterns, *Proc. Royal Society London B,* Vol. 244, pp. 21-26.

[15] Tomasi, C. and Birchfield, S. (1998) Depth Discontinuities by Pixel-to-Pixel Stereo, *Proc. IEEE Int. Conf. on Computer Vision.*

[16] Ullman, S. and Basri, R. (1991) Recognition by Linear Combinations of Models, *IEEE Trans. Patt. Anal. and Machine Intell.,* Vol. 13, No. 10, pp. 992-1006.

[17] Wolberg, G. (1990) *Digital Image Warping,* IEEE Computer Society Press, CA.

[18] Worthington P.L. and Hancock E.R. (1998) Increased Extent of Characteristic Views using Shape-from-Shading for Object Recognition, *Proc. BMVC98.*

[19] Worthington P.L. and Hancock E.R. (2000) View Synthesis using SFS, *Proc. Int. Conf. Pattern Recognition,* Vol. IV, pp. 110-113.

[20] Worthington P.L. and Hancock E.R. (2001) Object Recognition using SFS, *IEEE Trans. Patt. Anal. and Machine Intell.,* Vol. 22, pp. 535-542.

[21] Worthington, P.L. and Hancock, E.R. (1999) New Constraints on Data-Closeness and Needle Map Consistency for SFS, *IEEE Trans. Patt. Anal. and Machine Intell.,* Vol. 21, No. 12, pp. 1250-1267.

[22] Zhao, W.Y. and Chellappa, R. (2000) SFS-based View Synthesis for Robust Face Recognition, *Proc. IEEE Conf. on Automatic Face and Gesture Recog.,* pp. 285-292