Journal of Theoretical and Applied Information Technology

www.jatit.org

IMAGE-BASED MODELING: A REVIEW

Abas Md Said, Halabi Hasbullah, Baharum Baharudin
Computer & Information Sciences Dept., Universiti Teknologi PETRONAS, Malaysia
E-mail: abass@petronas.com.my, abasmdsaid@yahoo.com

ABSTRACT

We review work on image-based 3D modelling. 3D modelling entails the availability of complete 3D coordinates, and yet obtaining the coordinates is one of the most difficult tasks. Resorting to using 2D images appears a promising means to achieve this task and has been much reported. Prevalent methods are the shape from shading, photometry, stereopsis, photogrammetry and the video approach. Shape from shading takes minimal input (one image) but requires huge mathematical complexity. Photometry takes in at least twice as much input but produces more accurate results. Stereopsis takes in input of similar order but offers even better accuracy with less mathematical complexity. Photogrammetry uses even more input and produces good results in some types of applications. Video approach does away with assumptions in the previous methods and produces good results. However, stereopsis, photogrammetry and video approaches suffer from the correspondence problem.

Keywords: *Shape from shading, photometry, stereopsis, photogrammetry and shape from video.*

1 INTRODUCTION

In 3D applications, geometrical objects such as spheres or regular polyhedrons can be easily modelled. When irregular objects like those of human being and animals need to be incorporated, the task becomes more challenging and daunting. The main problem lies in determining the 3D coordinates. Some techniques have already been deployed to address this issue. Horn's seminal work to acquire 3D data from a single image (hence termed 'shape from shading') was notable (Horn, 1970). Decent results were obtained with specific assumptions (Bichsel & Pentland, 1992), (Dragnea & Angellopoulou, 2005), (Horn, 1970), (Horn & Brooks, 1985), (Oliensis, 1991), (Worthington, 2005), (Zhang et al., 1999). Other creative and innovative approaches to model 3D objects have been attempted and developed; among which are shape from stereopsis (Akimoto et al., 1993), (Shihong et al., 1999), shape from video (Brodski et al., 1999), (Strecha et al., 2003), photometric (Marlzbender et al., 2006), (Rushmeier et al., 1997), (Woodham, 1980) and photogrammetry (Grun et al., 2000).

Since all these works were based on 2D images, we now look at some of their basic relevant characteristics. A digital image is composed of a rectangular array of pixels (picture elements). A pixel holds the value of brightness at that particular discrete point usually using a combination of RGB (red, blue and green) colour space. Without loss of generality, a highly magnified section of an 8-by-8 pixel grey-scale digital image may be represented like the picture in Figure 1, where each small square represents a pixel. The light intensity value for each pixel that makes up the letter 'L', for example, may have a very low pixel value, while other pixels (white squares) have values one.

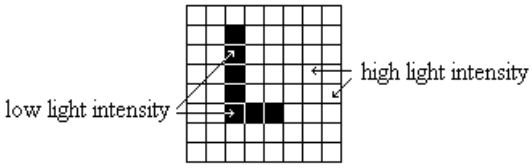


Figure 1: Letter 'L'.

We can use these brightness intensities to derive the depth (*z*-coordinates); the *x*- and *y*-coordinates can be arbitrarily determined. However, direct use of these pixel values will not give good results. The resulting intensities of the image are dependent on a few factors, namely, the surface colour, geometry, material and orientation of the object, and the surrounding light environment. The light environment effects may come largely from the direction of the source light,

the number of light sources and other perturbations like shadows and reflection.

2 SHAPE FROM SHADING

As a digital image is two-dimensional, we can easily assign the *x*- and *y*-coordinates using the image width and height, but we need a method to get the depth (*z*-coordinates) to create the 3D image. A rudimentary reconstruction of the 3D images based on the pixel values alone as the depth values will not adequately represent the image surface. This is due to the factors mentioned earlier. However, this piece of handy information (the pixel values) can be used to generate other valuable pieces of information.

Two key issues underlying the shape from shading (SFS) are determining the surface orientation of the object and generating depth once the orientation has been approximated. Most work in SFS has been based on the Lambertian model of shading. In the Lambertian model, a surface is assumed to be diffuse, which suggests that each surface point appears equally bright from all viewing direction. In such an environment, light intensity is dependent on the surface orientation and the direction of the light source. The surface orientation is typically represented by the normal vector **N**, i.e., a vector perpendicular to the surface, and the direction of light source, the vector **L**. This can be depicted as in Figure 2.

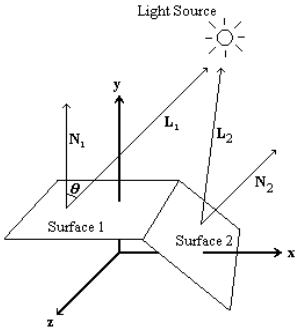


Figure 2: Surface orientation.

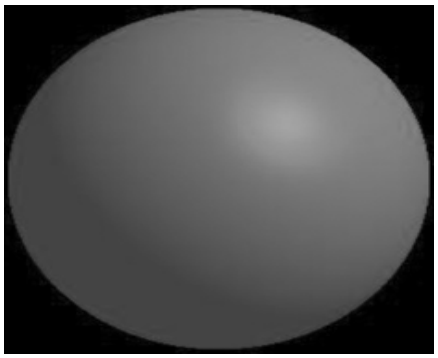
Based on the model, the resulting light intensity **I** on a surface (or at a particular pixel), for example, can be represented as follows:

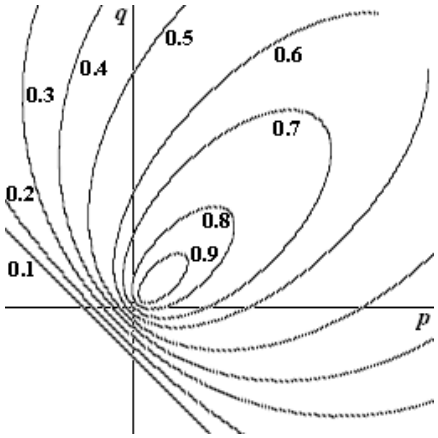
$\mathbf{I} = \mathbf{L_1} \cdot \mathbf{N_1}$ which is mathematically equivalent to $\mathbf{I} = |\mathbf{L_1}||\mathbf{N_1}| \cos\theta$.

This essentially means that the light intensity depends on the angle between the direction of the light source and the direction where the each planar surface patch is facing.

The brightness of each pixel in the image is related to the orientation of a corresponding patch on the surface of the object, other environmental factors ignored. This relationship between surface orientation and image brightness is fundamental to the SFS problem formulation. The shape of the object can be approximated once the orientation of each surface patch is known. To formulate the SFS problem, we assume the *z*-axis projects directly towards the viewpoint and *x*- and *y*- axes coincide with the axes of the image plane. We will assume that projection is orthographic (parallel) and the illumination source is at infinity.

The visible surface can be expressed by the explicit function $z(x,y)$. Instead of using **N**, the orientation of a surface point can be concisely represented by a pair (p,q) where $p = \partial z/\partial x$ (partial derivative of z with respect to x) and $q = \partial z/\partial y$ (partial derivative of z with respect to y), i.e., p and q are rates of change of the surface height in the x and y directions respectively. p and q values form the gradient space for the reflectance map. A reflectance map gives the pixel brightness as a function of the orientation of the scene surface in camera coordinates. In known case of a sphere, for example, a pixel with brightness 0.9 may be obtained possibly from any values of p and q on the ellipse in Figure 3.


a)



b)

Figure 3: (a) A sphere and (b) a possible reflectance map.

The relationship between the image and the surface gradient is given by the image irradiance equation, $\mathbf{E}(x,y) = \mathbf{R}(p,q)$, where **E** is the original image and **R** is the reflectance function which embodies both the effects of the imaging geometry and the reflectance properties of the surface. The SFS problem is to recover the surface function $z(x,y)$ by inverting the image irradiance equation or image-forming process. With some suitable rescaling, we may write **R** as a dot product of **L** and **N**, $\mathbf{R} = \mathbf{L} \cdot \mathbf{N}$, i.e., the predicted image intensities in terms of the surface orientation.

Following our definition of p and q, we can identify two vectors lying on an arbitrary tangent plane, namely $(1, 0, p)^T$ and $(0, 1, q)^T$ (where T stands for transpose). The cross product of these vectors gives us a normal vector $(-p, -q, 1)^T$, which can be normalized to $(-p, -q, 1)^T / \sqrt{(1 + p^2 + q^2)}$. The light source vector can be written as $\mathbf{L} = (l_1, l_2, l_3)^T$. For each pixel, if we write the observed image intensity as $i_{x,y}$ the equation yields

$$i_{x,y} = (- l_1 p - l_2 q + l_3) / \sqrt{(1 + p^2 + q^2)}.$$

If we simplify the condition by assuming the light direction as $(0, 0, 1)^T$, i.e., along the z-axis (though not always the case), and taking the brightest intensity at a particular pixel, for example, we get $1 = 1 / \sqrt{(1 + p^2 + q^2)}$. This system of equations is clearly under-constrained. Thus, framing the problem mathematically leads to having more unknowns than constraints and many researchers consider this problem ill-posed [3, 27]. To get around the problem of ill-posedness, the following assumptions are usually made: i) the surface is uniform in its reflecting properties, ii) the light sources are far away so that the irradiance of different parts of the scene will be approximately the same and the incident direction may be taken as constant, and iii) the viewer is far away, so that the direction to the viewer will be the same for all points in the scene. The next few subsections discuss approaches attempted to solve the irradiance equation.

2.1 Variational Method

To find a surface that satisfies the image irradiance equation as well as possible, many algorithms minimize the brightness error between the predicted image of the surface $\mathbf{R}(p,q)$) and the observed image $\mathbf{E}(x,y)$:

$$\mathbf{Q} = \sum [\mathbf{E}(x,y) - \mathbf{R}(p,q)]^2$$

i.e., instead of looking for an exact solution to the problem, we allow some deviations between the image brightness and the reflectance map. This problem is commonly transformed into a partial differential equation problem, with an additional smoothness constraint to control and stabilize the smoothness of the solution:

$$\iint_\Omega (\mathbf{E}(x,y) - \mathbf{R}(f(x,y),g(x,y)))^2 + \lambda (f_x^2 + f_y^2 + g_x^2 + g_y^2)\,dxdy \quad [14].$$

It is then solved by looking for a function that minimizes the functionals, in this case using the calculus of variation. Critics usually point out that algorithms that use a smoothness constraint typically give unrealistically smooth results (surface) (Gultekin & Gokmen, 1998). In addition, convergence of solution tends to be very slow; some might take a thousand iterations (Dragnea & Angelopoulou, 2005).

2.2 Propagation Method

In this method, the surface is developed by starting from points of known orientation or height values (Bichsel & Pentland, 1992), (Kimmel et al., 1995). The process is performed iteratively along a certain path in the image. For example, if the height value Z_0 at the point (x_0,y_0) is known in advance, all Z values can be determined along the path by calculating a line integral such as $\boldsymbol{Z(t) = Z_0 + \int_0^t \frac{dZ}{dt} dt}$. The drawback of this method is that the locality of calculations causes a high dependency on data accuracy and the propagation of height increments along paths also

means the propagation of errors (Klette et al., 1998).

2.3 Local Method

This is a method where the shape at an image point calculated by using the image irradiances only in a small neighbourhood of that image point (Lee & Rosenfeld, 1985), (Pentland, 1984). The advantage is time efficiency because the determination of the shape is locally constrained. Furthermore, the approach is non-iterative. The idea is to approximate surface orientation as a sphere to resolve ambiguities associated with an image irradiance local to a point. It can only determine surface orientation; other methods, such as integration, have to be used to arrive at the height map.

Apparently the application of the partial differential equations theory to the SFS problem has been hampered by several difficulties. The first type arises from the simplification introduced in the modelling: orthographic cameras looking at Lambertian objects with a single point light source at infinity is the set of usual assumptions (Horn & Brooks, 1985), (Zhang et al., 1999). The second type is mathematical: characterizing the solutions of the corresponding partial differential equations has turned out to be a very difficult problem (Prados et al., 2002). The third type is algorithmic: assuming that the existence of a solution has been proved, coming up with provably convergent numerical schemes has turned out to be quite involved (Durou & Mitre, 1996).

3 PHOTOMETRIC STEREO

Photometric stereo was pioneered by (Woodham, 1980). In the photometric stereo method, multiple non-collinear light sources are used to obtain different images of the surface, with the same position in each image relating to the same surface point. The orientation of each surface patch is calculated directly from the intensities of the corresponding points in the image. Technically two light sources should be adequate, but having more than three can compensate for problematic samples such as shadows and specular regions. This is done by removing the unwanted sets of samples at the image location (Marlzbender et al., 2006).

A set-up similar to the one used by (Rushmeier et al., 1984), for example, is shown in **Figure 4**. Once the orientation of every surface patch has been found, a depth map that explains these orientations is computed using, for example, the calculus of variations and an iterative technique.

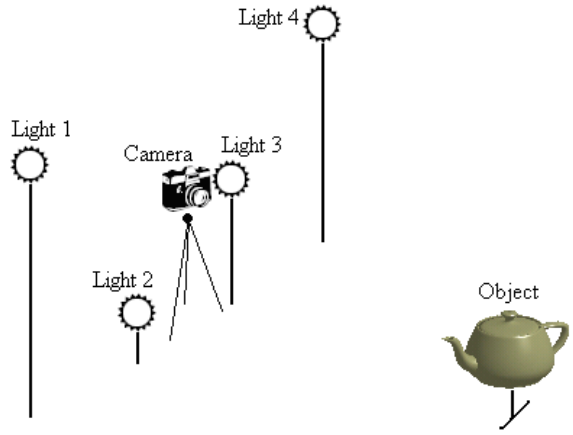


Figure 4: A lighting set-up for photometry.

Conceptually, the method is similar to shape from shading, following the Lambertian model, except that the gradients are more accurately determined. Based on $\mathbf{E}(x,y) = \mathbf{R}(p,q)$, but writing the surface orientation in terms of the normal vectors rather than gradients, for each point p, we obtain

$$\rho \begin{bmatrix} l_{1,x} & l_{1,y} & l_{1,z} \\ l_{2,x} & l_{2,y} & l_{2,z} \\ l_{3,x} & l_{3,y} & l_{3,z} \end{bmatrix} \begin{bmatrix} n_{p,x} \\ n_{p,y} \\ n_{p,z} \end{bmatrix} = \alpha \begin{bmatrix} E_{p,1} \\ E_{p,2} \\ E_{p,3} \end{bmatrix}$$

ρ = reflectance (albedo) of surface
α = constant to incorporate light source brightness, camera sensitivity, etc.,
l_{ij} = light source vector components using light source i = 1, 2 or 3 at $j = x$, y, or z-coordinate
$n_{p,j} = j^{th}$ normal vector components at point p, with $j = x$, y, or z-coordinate, and
$E_{p,j}$ = image intensity at point p corresponding to light source j = 1, 2 or 3.

Good results have been reported by this method (Bertesaghi et al., 2005), (Georghiades, 2003), (Marlzbender et al., 2006).

4 STEREOPSIS

The depth can also be approximated using the ideas in stereo vision. By taking photographs from two different locations (left and right), the "lines of sight" can be developed from each camera to points on the object. These lines of sight are intersected to calculate the 3D coordinates of the points of interest, using triangulation.

However, when the same scene is taken from two different views, the orientations of the images are different (Figure 5). This can be dealt with by using epipolar geometry. The left and right images are projected on the scanline plane in such a way that the epipolar lines become parallel to the scanline. The results of this projection are slightly warped (distorted) images (Figure 6). Corresponding points (objects) in both images have to be identified along each epipolar line. This is often difficult and commonly termed the correspondence problem in the computer vision circle. Having identified the corresponding points, depth can be calculated by triangulation, i.e., using similar triangles.

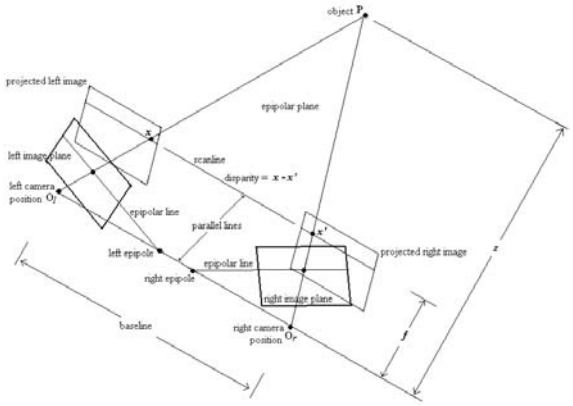


Figure 5: Epipolar geometry.

The main procedures can be sketched as follows:

Calibrate camera (camera resection)
Rectify images
for each epipolar line
for each pixel in the left image
compare with every pixel on same epipolar line in right image
pick pixel most matching in the right (use windows)
Estimate depth using similar triangles, $x - x'$ / baseline = f / z, i.e., $z = f \cdot$ baseline / $(x - x')$.

Algorithms have been developed to handle the correspondence problem. Some use feature-based approach by matching edges and producing sparse depth maps while others adopt correlation technique by matching all pixels in the entire images and producing dense depth maps. Although good results can be achieved with stereo techniques, the excessively long computation time needed to match stereo images is still the main obstacle on the way to their practical applications. General purpose computers are not fast enough to meet real-time requirements because of the algorithmic complexity of stereo vision techniques.

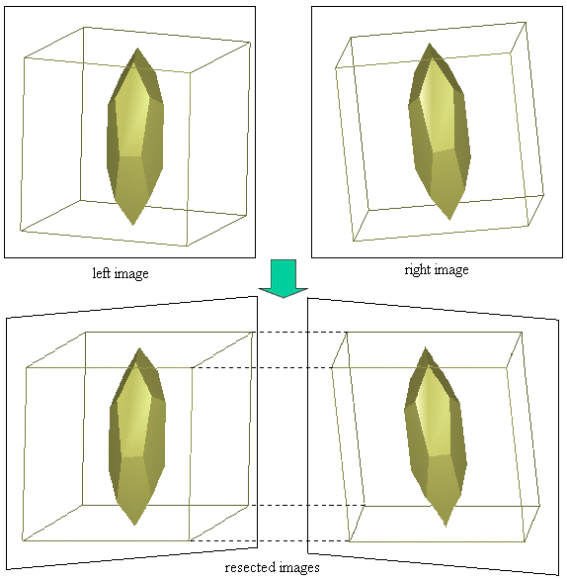


Figure 6: Images after resection.

5 PHOTOGRAMMETRY

As its name suggests, photogrammetry is a 3D coordinate measuring technique using photogram, although photographs are more commonly used now as the fundamental medium for measurement. The technique has been used traditionally in surveying and mapping for the reconstruction of terrain, natural targets and buildings in stereovision. The advances in electronics technology have made it applied in various areas; archaeology, computer vision and computer graphics (Debevec et al., 1996), (Grun et al., 2004), (Remondino, 2006), (Xia & Zhu, 2005). Depth can be calculated by applying triangulation on usually two photographs. Using a digital camera with known characteristic (e.g., lens focal length, imager size and number of pixels), a minimum of two pictures of an object are needed.

If we can indicate the same three object points in the two images and we can indicate a known dimension we can determine other 3D points in the images.

These techniques mostly require very precise calibration and there is almost no automation. The detailed acquisition of models is therefore very time consuming. The recent decade saw researchers trying to both reduce the requirements for calibration with the goal to automatically extract a realistic 3D model by freely moving a camera around an object. An early approach was proposed by (Tomasi and Kanade, 1992). They used an affine factorization method to extract 3D data from image sequences, assuming orthographic projection. Another type of system starts from an approximate 3D model and camera poses and refines the model based on images, e.g., *Facade* proposed by (Debevec et al., 1996). The advantage is that fewer images are required. On the other hand a preliminary model must be available and the geometry should not be too complex.

The basic idea is simple: when a picture is taken, the 3D world is projected in perspective onto a flat 2D image plane. As a result, a feature (for example, the top of a flagpole) seen at a particular location in an image can actually lie anywhere along a particular ray beginning at the camera centre and extending out to infinity. This ambiguity can be resolved if the same feature is seen in two different photographs, which constrains the feature to lie on the intersection of the two 'corresponding rays'. This process is known as *triangulation.* Using triangulation, any feature seen in at least two photographs taken from known locations can be localized in 3D. In fact, with a sufficient number of corresponding points, it is mathematically possible to solve for unknown camera positions as well.

The application of (close-range) photogrammetry in computer graphics for 3D surface reconstruction often involves a sequence of images taken with off-the-shelf consumer cameras. The user acquires the images by freely moving the camera around the object. Neither the camera motion nor the camera settings have to be known, in contrast to the requirements in many other computer vision algorithms. This approach has been developed over the last few years. The stereopsis method is still applied here, minus the so-called 'calibrations'. The performance analysis showed that very dense depth maps with fill rates of over 90 % and a relative depth error of 0.1% can be measured with off-the-shelf cameras even in unrestricted outdoor environments such as an archaeological site (Pollefeys et al., 1999).

6 SHAPE FROM VIDEO

Obtaining 3D structure from video is a problem with a long tradition in computer vision, e.g., depth from stereo and depth from motion. Although it would seem video-based reconstruction could benefit from stereopsis, one of the most successful approaches has been by the factorization method, first proposed by Tomasi and Kanade for rigid shape (Tomasi & Kanade, 1992). With two images of reasonable disparity, stereopsis is well-posed; on the other hand, a video sequence consists of many frames, implying over-constrainedness. The factorization method handles over-constrainedness by means of least-squares. The underlying concept is frequently termed structure from motion. First it applies the rank constraint to factorize a set of feature locations tracked across the entire sequence. Then it uses the orthonormality constraints on the rotation matrices to recover the scene structure and camera rotations. Proven fast and stable, this approach works under the orthographic projection model.

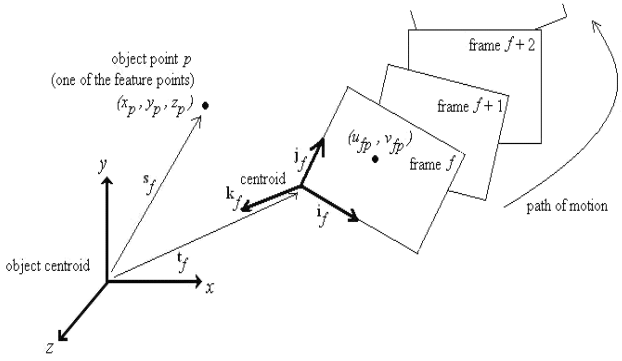


Figure 7: Frames and a feature point.

The factorization method is elegantly formulated in terms of matrices. Given an image stream, and assuming that we have tracked P feature points over F frames, we denote the trajectories of image coordinates by (u_{fp}, v_{fp}) such that $f = 1, \ldots, F$, and $p = 1, \ldots, P$ (Figure 7). We write $\widetilde{\mathbf{W}}$, the registered measurement matrix, as an augmented matrix of $\widetilde{\mathbf{U}}$ and $\widetilde{\mathbf{V}}$, i.e., $\widetilde{W} = \begin{bmatrix} \widetilde{U} \\ \widetilde{V} \end{bmatrix}$, storing all the relevant coordinates from the frames. Then, the entries of $\widetilde{\mathbf{U}}$ and $\widetilde{\mathbf{V}}$

will be $\widetilde{\mathbf{U}} = \begin{bmatrix} \widetilde{\mathbf{u}}_{11} & \cdots & \widetilde{\mathbf{u}}_{1\mathbf{P}} \\ \vdots & \cdots & \vdots \\ \widetilde{\mathbf{u}}_{\mathbf{F}1} & \cdots & \widetilde{\mathbf{u}}_{\mathbf{FP}} \end{bmatrix}$ and $\widetilde{\mathbf{V}} = \begin{bmatrix} \widetilde{\mathbf{v}}_{11} & \cdots & \widetilde{\mathbf{v}}_{1\mathbf{P}} \\ \vdots & \cdots & \vdots \\ \widetilde{\mathbf{v}}_{\mathbf{F}1} & \cdots & \widetilde{\mathbf{v}}_{\mathbf{FP}} \end{bmatrix}$, where

$$\widetilde{u}_{fp} = u_{fp} - \frac{1}{P}\sum_{p=1}^{P} u_{fp} \quad \text{and}$$

$\widetilde{v}_{fp} = v_{fp} - \frac{1}{P}\sum_{p=1}^{P} v_{fp}$.

Assuming orthographic projection, we can decompose $\widetilde{\mathbf{W}}$ as the product of two matrices **R** and **S**, $\widetilde{\mathbf{W}}$ = **RS**; the rows of **R** represent the orientations of the horizontal and vertical camera reference axes throughout the stream, while the columns of **S** are the coordinates of the P feature points with respect to their centroid. The decomposition of $\widetilde{\mathbf{W}}$ into **R** and **S** is carried out by the singular value decomposition into $\widetilde{\mathbf{W}} = \mathbf{O}_1 \Sigma \mathbf{O}_2$. Further, by partitioning, this can be written as $\widetilde{\mathbf{W}} = \mathbf{O}_1^{'}[\Sigma]^{1/2}[\Sigma]^{1/2}\mathbf{O}_2^{'}$ where $O_1^{'}[\Sigma]^{1/2}$ can be written as $\hat{\mathbf{R}}$ and $[\Sigma]^{1/2}O_2^{'}$ as $\hat{\mathbf{S}}$ (i.e., $\hat{\mathbf{R}} = \mathbf{O}_1^{'}[\Sigma]^{1/2}$ and $\hat{\mathbf{S}} = [\Sigma]^{1/2}\mathbf{O}_2^{'}$). In the absence of noise, $\hat{\mathbf{R}}$ is in fact a linear transformation of the true rotation matrix **R** and $\hat{\mathbf{S}}$ is a linear transformation of the true shape matrix **S**. Then, **R** and **S** can be written as $\mathbf{R} = \hat{\mathbf{R}}\mathbf{Q}$ and $\mathbf{S} = \mathbf{Q}^{-1}\hat{\mathbf{S}}$ as there exists such a 3 by 3 matrix **Q** and thus, the shape can be recovered from the matrix **S**.

Poelman and Kanade extended it to work under the weak perspective and para-perspective projection models (Poelman & Kanade, 1997). Triggs generalized the factorization method to the recovery of scene geometry and camera motion under the perspective projection model (Triggs, 1996). These methods work for static scenes. However, given a long video sequence it is often not practical to work with all video frames. In addition, to allow for effective outlier rejection and motion estimation it is necessary to have a baseline between frames. For this purpose, Repco and Pollefeys proposed a key-frame selection procedure based on a robust model selection criterion (Repco & Pollefeys, 2005). Their approach guarantees that the camera motion can be estimated reliably by analyzing the feature correspondences between three consecutive views.

There are a number of approaches that try to get a 3D model of the scene from uncalibrated images automatically. The fully automated procedure widely reported in the vision community starts with a sequence of images taken with an uncalibrated camera (Mayer, 2003), (Nister, 2001), (Pollefeys et al., 1999). The system then extract interest points, sequentially match them across the view-pairs and compute the camera parameters as well as the 3D coordinates of the matched points using robust techniques. This is done in a projective geometry framework and is usually followed by a bundle adjustment. A self-calibration, to compute the interior camera parameters, is afterwards performed in order to obtain a metric reconstruction, up to a scale, from the projective one. The 3D surface model is then automatically generated by means of dense depth maps on image pairs. The key to the success of these automated approaches is the very short interval between consecutive images. Some approaches have been presented for the registration of widely separated views but their reliability and applicability for automated image-based modelling of complex objects are still not satisfactory, as they yield mainly a sparse set of matched feature points. Dense matching results under wide baseline conditions were instead reported in (Strecha et al., 2003). Photogrammetry started more than 150 years ago. As its name implies, it is a 3D coordinate measuring technique using photogram, although photographs are more commonly used now as the fundamental medium for measurement. The technique has been used traditionally in surveying and mapping for the reconstruction of terrain, natural targets and buildings

7 CONCLUDING REMARKS

We have reviewed prevalent approaches to image-based 3D modelling. We can now summarise that shape from shading approach is the least on equipment requirements but at the price of accuracy and mathematical complexity, and in many cases the resulting surfaces are unreasonably smooth. Photometry is an improvement to SFS only in terms of achieving better accuracy on the gradients; but both eventually need to *estimate* the 3D surfaces from the gradients.

Stereopsis provides better accuracy but only with the availability of some values of environment and equipment parameters, e.g., the focal length of the camera lens and the distance between the cameras. Close-range photogrammetry essentially makes use of the stereopsis concepts of calculation but with more images. It has been successfully applied to modelling archaeological and architectural objects, but there are no proven results from irregular surfaces like human faces. In both methods, the problems lie in the correspondence of feature points in the images and making essential equipment calibrations.

The video approach renders the assumptions in all previous methods irrelevant but it still has to address the correspondence problem. Nevertheless, good results have been shown when the correspondence problem was correctly addressed.

ACKNOWLEDGEMENTS

We would like to thank Universiti Teknologi PETRONAS for supporting this work in all aspects and Prof. Kuniaki Uehara of Kobe University for playing host for two months at the Graduate School of Science & Technology.

REFERENCES

[1]. Akimoto, T., Y. Suenaga, and R. Wallace, 1993. Automatic creation of 3D facial models. IEEE Computer Graphics & Applications, 13(5): 16 - 22.

[2]. Bertesaghi, A., G. Sapiro, T. Malzbender, and D. Gelb, 2005. Three-dimensional shape rendering from multiple images. Graphics Models, 67(2005): 332 - 346.

[3]. Besl, P. J. and R. C. Jain, 1985. Three-dimensional object recognition. Computing Surveys, 17(1): 75 - 144.

[4]. Bichsel, M. and A. P. Pentland, 1992. Simple algorithm for shape from shading. In: Proc. of IEEE Computer Vision and Pattern Recognition, pp. 459 - 465.

[5]. Brodski, T., C. Fermuller and Y. Aloimonos, 1999. Shape from video. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2: 1 - 151.

[6]. Debevec, P. E., C. J. Taylor and J. Malik, 1996. Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach. In: Proc. of the 23rd Annual Conf. on Computer Graphics and Interactive Techniques SIGGRAPH '96.

[7]. Dragnea, V. and E. Angelopoulou, 2005. Direct shape from isophotes. In: Proc. of the ISPRS Workshop BenCOS 2005.

[8]. Durou, J.-D. and H. Maître, 1996. On convergence in the methods of Strat and Smith for shape from shading. Int. Journal of Computer Vision, 17(3): 273 - 289.

[9]. Georghiades, A. S., (2003). Measurement and colour matching: Recovering 3D shape and reflectance from a small number of photographs. In: Proc. of 13th Eurographics Workshop on Rendering, pp. 230 - 240.

[10]. Grün, A., 2000. Semi-automated approaches to site recording and modelling. Int. Archives of Photogrammetry and Remote Sensing, 33(5/1): 309 - 318.

[11]. Grün, A., F. Remondino and L. Zhang, 2004. Photogrammetric reconstruction of the Great Buddha of Bamiyan, Afghanistan. Photogrammetric Record, 19(107): 177 -199.

[12]. Gültekin, A. & M. Gökmen, 1998. Adaptive shape from shading. Elektrik, 6(2): 61-73.

[13]. Horn, B. K. P., 1970. Shape from shading: A method for obtaining the shape of a smooth opaque object from one view. Technical Report No. 232, AI Lab., MIT.

[14]. Horn, B. K. P. and M. J. Brooks, 1985. The variational approach to shape from shading. AI Memo 813, AI Lab., MIT.

[15]. Horn, B. K. P. and M. J. Brooks, 1989. Shape from Shading. The MIT Press, 1989.

[16]. Kimmel, R, K. Siddiqi, B. Benjamin, B. B. Kimia and A. M. Bruckstein, 1995. Shape from shading: Level set propagation and viscosity solutions. Int. Journal of Computer Vision, 16: 107-133.

[17]. Klette, R., K. R. Schluns and A. Koschan, 1998. Three-dimensional data from images, Singapore, Springer-Verlag.

[18]. Lee, C.-H. and A. Rosenfeld, 1985. Improved methods of estimating shape from shading using light source coordinate system. Art. Intelligence, 26(1985): 125 - 143.

[19]. Malzbender, T., B. Wilburn, D. Gelb and B. Ambrisco, 2006. Surface enhancement using real-time photometric stereo and reflectance transformation. Eurographics Workshop on Rendering 2006.

[20]. Mayer, H., 2003. Robust orientation, calibration and disparity estimation of Image Triplets. In: 25th DAGM Pattern Recognition Symposium.

[21]. Nister, D., 2001. Automatic dense reconstruction from uncalibrated video sequences. PhD Thesis, Royal Institute of Technology, Stockholm.

[22]. Oliensis, J., 1991. Uniqueness in shape from shading. Int. Journal of Computer Vision, 6(2): 75 -104.

[23]. Pentland, P., 1984. Local shading analysis. IEEE Trans. PAMI, 6(2): 170 - 187.

[24]. Poelman, C. and Kanade, T., 1997. A para-perspective factorization method for shape and motion recovery. IEEE Trans. Pattern Analysis and Machine Intelligence, 19(3): 206 - 218.

[25]. Pollefeys, M., R. Koch, M. Vergauwen and L. Van Gool, 1999. An automatic method for acquiring 3D models from photographs. In: Proc. of ISPRS Int. Workshop on Photogrammetric Measurements, Obj. Modelling and Doc. in Architect. and Industry.

[26]. Prados, E., O. Faugeras and E. Rouy, 2002. Shape from shading and viscosity solutions. In: Proc. of European Conf. On Computer Vision.

[27]. Prados, E., F. Camilli and O. Faugeras, 2006. A unifying and rigorous shape from shading method adapted to realistic data and applications. Journal of Mathematical Imaging and Vision, 25(3): 307 - 328.

[28]. Remondino, F., 2004. 3-D reconstruction of static human body shape from image sequence. Computer Vision and Image Understanding, 93(1): 65 - 85.

[29]. Remondino, F., 2006. Image-based modeling for object and human reconstruction. PhD Thesis, Swiss Federal Institute of Technology, Zurich

[30]. Remondino, F. and S. El-Hakim, 2006. Image-based 3D modelling: A review. Photogrammetric Record, 21(115): 269 - 291.

[31]. Repco, J. and M. Pollefeys, 2005. 3D models from extended uncalibrated video sequences. In: Proc. of 5th Int. Conf. on 3D Dig. Imaging and Modelling, pp. 150 - 157.

[32]. Rushmeier, H., G. Taubin and A. Gueziec, 1997. Applying shape from lighting variation to bump map capture. In: Proc. of Eurographics Work. on Rendering, 35 - 44.

[33]. Shihong, L., Y. Sumi, M. Kawade and F. Tomita, 1999. Building 3D facial models and detecting face pose in 3D space. In: Proc. of 2nd Int. Conf. on 3D Digital Imaging and Modelling, 1999.

[34]. Strecha, C., F. Verbiest, M. Vergauwen and L. Van Gool, 2003. Shape from video vs. still images. Optical 3D Measurement Techniques, Vol. 2, pp. 168-175.

[35]. Tomasi, C. and T. Kanade, 1992. Shape and motion from image streams under orthography: A factorization method. Int. Journal of Computer Vision, 9(2):137 - 154.

[36]. Triggs, B., 1996. Factorization methods for projective structure and motion. In: Proc. of Int. Conf. of Computer Vision and Pattern Recognition,1996.

[37]. Woodham, R. J., 1980. Photometric method for determining surface orientation from multiple images. Optical Engineering, 19(1): 139 - 144.

[38]. Worthington , P. L., 2005. Reillumination-driven shape from shading. Computer Vision and Image Understanding, 98(2005): 326 - 344.

[39]. Xia, S. and Y. Zhu, 2005. 3D simulation and reconstruction of large-scale ancient architecture with techniques of photogrammetry and computer science. In: CIPA 2005 XX International Symposium.

[40]. Zhang, R., P.-S. Tsai, J. E. Cryer and M. Shah, 1999. Shape from shading: A survey. Trans. on Pattern Analysis and Machine Intelligence, 21(8): 690 - 706.