AN INTEGRAL PROJECTION APPROACH TO 3D RIGID BODY TRANSFORMATIONS

Stefan Lehmann, I. Vaughan L. Clarkson, and Peter J. Kootsookos

Intelligent Real-Time Imaging and Sensing Group
School of Information Technology and Electrical Engineering
The University of Queensland
4072 AUSTRALIA
(lehmann,v.clarkson,kootsoop)@itee.uq.edu.au

ABSTRACT

The analysis of 3D rigid body transformations based on camera images is of great significance in many research areas. Classical methods to recover 3D object information often use a set of cameras that cover a static scene from different view angles. Key areas in this context include Structure from Motion, Motion from Structure and Camera Calibration. This paper introduces a new approach to analyse 3D rigid body transformations which we call integral projection. In this model, we are able to use frequency-domain information to estimate parameters of the transformation. Simulations are presented which demonstrate our initial successes.

1. INTRODUCTION

One of the most challenging research areas in computer vision is to gain 3D information about an object from camera images of this object. In the case of a static object, at least two cameras with different view angles are required to compute depth maps that enable back-projection of the 2D pixels, *i.e.*, the transformation of a 2D pixel into the corresponding 3D voxel of the object [1]. On the other hand, for dynamic scenes, the use of one camera only allows the acquisition of 3D information, provided that at least two frames of an image sequence from that camera are being processed. Luong and Faugeras, amongst many others, have shown how 3D structure and motion information can be estimated based on point correspondences [2].

Gaining structural information about the object or scene based on rigid body transformations of a 3D object or scene is known as *Structure from Motion*. A variety of approaches have been proposed in this context [3, 4, 5]. An overview of different methods can be found in [6]. On the other hand, the aim of *Motion from Structure* is to extract 3D motion parameters from camera images. In the dynamic case, a sequence of images from one camera can be analysed to determine the 3D motion parameters of a moving object. In the static case, the images of a fixed scene from multiple cameras can be used for camera calibration.

Finding the depth information of the image pixels relies on an accurate determination of rigid body transformations, *i.e.*, the relative rotation and translation of the 3D object with respect to the camera. The camera model determines the relationship between the 3D voxels of the object and the corresponding 2D pixels in the image plane. In computer vision, the most commonly used projection models are either parallel or perspective projections.

Transformations in the space domain that are composed of rotations and translations correspond to pure rotations of the spectral magnitudes in the frequency domain [7]. A spatial translation yields a phase shift in the frequency domain. However, common projection techniques such as parallel projection are non-linear transformations between the 3D voxel model and the 2D image model. As a consequence, the transformation between the two Fourier spectra of the images that result from the projections of the original and the transformed object respectively is not straightforward.

We propose an integral projection model that is a linear operation and establishes a straightforward relationship between the two image spectra. There is a correspondence between parallel and integral projection which will be discussed in Section 2.1.

Section 2 introduces our model and shows its relevance for rigid body transformations. An algorithm for estimating transformation parameters is presented in Section 3. Initial experimental results are shown in Section 4. We conclude with an outlook on future research.

2. INTEGRAL PROJECTION

2.1. Concept and relationship to parallel projection

We will illustrate the integral projection scheme by projecting a simple 2D object into a 1D projection function. The integral projection model determines the 1D projection values by integrating the 2D object along lines that run parallel to the view axis. This model comprises two simplifications: use of integration to perform the projection and use of parallel, rather than fan or perspective, projection.

As we pointed out in Section 1, there is a duality between *Structure from Motion* and *Motion from Structure*. Recording static scenes with multiple identical cameras from different viewpoints is equivalent to recording dynamic scenes or objects with one static camera. Without loss of generality, the following considerations are based on the assumption that we have one static camera that is recording a moving object.

Suppose the 2D object in its original position is represented by $f_2(y, z)$. After an object transformation has taken place, the object is represented by $f_2'(y, z)$. Integrating $f_2(y, z)$ and $f_2'(y, z)$ along lines that are parallel to the view axis results in the 1D integral projection values. Figures 1 and 2 depict this situation. In both figures, the view axis is the z-axis of the object coordinate system. The integral projections are 1D functions that reflect the integration results as intensity values $I(y)$ and $I'(y)$ respectively.

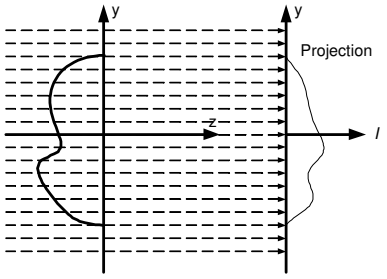

Fig. 1. Integral projection of an object before transformation

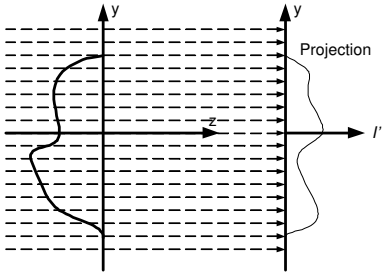

Fig. 2. Integral projection of an object after transformation

For the following considerations we will assume that the objects in Figures 1 and 2 are delta-surface objects, i.e. $f_2(y, z)$ and $f_2'(y, z)$ are non-zero only in points (y, z) that belong to the object surface. We further assume that $f_2(y, z)$ and $f_2'(y, z)$ are non-zero only for surface points that are in the field of the view of a camera that is described by a parallel camera model. These surface points are depicted in Figures 1 and 2. Under these assumptions, integral projection is identical to parallel projection.

For general object geometries, object faces might become obscured or revealed after object transformations. Therefore, the delta surface functions $f_2(y, z)$ before and $f_2'(y, z)$ after an object transformation are generally not only transformed versions but differently shaped as shown in Figures 1 and 2. There will consequently be deviations between applying an integral projection or a parallel projection to the transformed surface function $f_2(y, z)$, even though both projection models have identical results for the original surface function $f_2(y, z)$.

However, the object faces that become obscured or revealed after transformations can be expected to be relatively small if the camera frame rate is high and the object transformation between the frames is relatively small. Therefore, the strong correspondence between parallel and integral projection is retained. Perspective projections can be approximated by parallel projections if the distance of the object to the camera is relatively large compared to the focal length.

2.2. Mathematical model

Suppose the 3D scene can be represented by a function $f_3(x, y, z)$. The integral projection model determines the 2D pixel values by integrating $f_3(x, y, z)$ along lines that are running parallel to the view axis. We consider an ideal image plane with infinite extent and describe continuous images rather than digital images in our mathematical model. Using our integral projection approach and assuming that the z-axis of our object coordinate system is aligned with the view axis of the camera, the corresponding 2D image data will be

$$f_2(x, y) = \int_{\mathbb{R}} f_3(x, y, z)\, dz. \tag{1}$$

The Fourier spectra of $f_2(x, y)$ and $f_3(x, y, z)$ can be denoted as

$$F_2(\xi, \eta) = \int_{\mathbb{R}^2} f_2(x, y) e^{-j(\xi x + \eta y)}\, dx\, dy \tag{2}$$

$$F_3(\xi, \eta, \zeta) = \int_{\mathbb{R}^3} f_3(x, y, z) e^{-j(\xi x + \eta y + \zeta z)}\, dx\, dy\, dz. \tag{3}$$

Thus, the relationship between the Fourier spectra $F_2(\xi, \eta)$ and $F_3(\xi, \eta, \zeta)$ is

$$\begin{aligned} F_2(\xi, \eta) &= \int_{\mathbb{R}^3} f_3(x, y, z) e^{-j(\xi x + \eta y)}\, dx\, dy\, dz \\ &= F_3(\xi, \eta, 0). \end{aligned} \tag{4}$$

Equation 4 is known as the projection-slice theorem [8]. An important application of this theorem is found in tomography [9] where it is deployed for recovering x-ray images. As (4) shows, the spectrum of the image data is the $\xi, \eta-$plane (where $\zeta = 0$) of the corresponding spectrum of the 3D object data. What does this relationship mean for 3D rigid body transformations? This will be discussed in Section 2.3.

2.3. Effect of 3D rigid body transformations

For the following considerations, we introduce the vectors

$$\Delta = \left(\xi, \eta, \zeta\right)^T \tag{5}$$

and

$$\Lambda = \left(x_0, y_0, z_0\right)^T \tag{6}$$

where ξ, η and ζ represent the 3D frequency components of $F_3(\xi, \eta, \zeta)$ in (3) and x_0, y_0 and z_0 are the translational components of the transformation with respect to the x, y, z coordinate axes. Let us now consider that we have the two 2D integral projection images of a rigid body that is subject to the following linear transformation

$$P' = RP + \Lambda \tag{7}$$

with R representing the rotation matrix, $P = (x, y, z)^T$ a point of the object and $P' = (x', y', z')^T$ the corresponding transformed point. After the transformation, the 3D spectrum that corresponds to the transformed object will be

$$F_3'(\Delta) = e^{-j(\Lambda^T \Delta)} F_3(R^T \Delta). \tag{8}$$

According to (4), the 2D spectrum of the image is the $\zeta = 0$ plane of the 3D spectrum of the object. Therefore, the two spectra of the projection images before and after the transformation of the object will show matching lines that run through the origin of the coordinate systems of the spectra. Figure 3 depicts this situation:

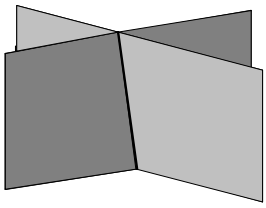

Fig. 3. Matching lines

The magnitudes of the two spectra along these lines will be identical. The phases will show an offset which depends upon the translational component of the transformation. Suppose we could detect matching lines in the 2D Fourier spectra. What does this tell us about the rigid body transformation that has taken place?

2.4. Analysis of the transformation parameters

Let us assume that (ξ, η) and (ξ', η') are corresponding frequency locations along the matching lines of the spectrum $F_2(\xi, \eta)$ of the first and the spectrum $F_2'(\xi', \eta')$ of the second image respectively. Equation 4 yields the following relationship

$$F_2'(\xi', \eta') = F_3'(\xi', \eta', 0) \tag{9}$$

and finally with (8)

$$F_2'(\xi', \eta') = e^{j(\xi' x_0 + \eta' y_0)} F_2(\xi, \eta). \tag{10}$$

Detecting the matching lines in the two 2D spectra would yield two types of information. Firstly, by recovering the phase shift factor $e^{j(\xi' x_0 + \eta' y_0)}$ in equation (10), we gain information about the translational components x_0 and y_0 of the object transformation. Even though x_0 and y_0 can not be isolated from the phase shift factor, we can reveal information about their relationship. Secondly, additional information is contained in the angle pair (α, α') of the matching lines with respect to the ξ- and ξ'-axes of the frequency spectra $F_2(\xi, \eta)$ and $F_2'(\xi', \eta')$ respectively. What does this angle pair tell us about the 3D rigid body transformation?

To answer this question, we will examine how (α, α') depend on the rotation of the object about an azimuth angle ϕ and an elevation angle θ around the origin of the object coordinate system. Defining our xyz-coordinate system such that the elevation θ denotes the rotation of our object about the $x-$axis and the azimuth ϕ about the $y-$axis, the corresponding rotation matrix will be

$$R = \begin{pmatrix} \cos\phi & 0 & -\sin\phi \\ \sin\phi\,\sin\theta & \cos\theta & \cos\phi\,\sin\theta \\ \sin\phi\,\cos\theta & -\sin\theta & \cos\phi\,\cos\theta \end{pmatrix} \tag{11}$$

Rotation matrix R transforms each voxel $(x, y, z)^T$ of the object into a corresponding voxel $(x', y', z')^T$ according to (7). Equation 8 shows that R also establishes the transformation between corresponding frequency indices in the 3D Fourier spaces of the original and the transformed object. With respect to (4) and (9), the transformation of the frequency pair ξ and η of the projection of the object into the corresponding matching frequencies ξ', η' of the projection of the transformed object is described by

$$(\xi', \eta', 0)^T = R \cdot (\xi, \eta, 0)^T \tag{12}$$

This yields the relationship between ξ and η along the matching line depending on the angles ϕ and θ. Thus, the angle α of the matching line with respect to the $\xi-$axis of the spectrum can be found. Similarly, the angle α' can be determined from

$$(\xi, \eta, 0)^T = R^T \cdot (\xi', \eta', 0)^T. \tag{13}$$

The results for α and α' are

$$\alpha = \arctan\left(\frac{\sin\phi\,\cos\theta}{\sin\theta}\right) \tag{14}$$

$$\alpha' = \arctan\left(\frac{\sin\phi}{\cos\phi\,\sin\theta}\right) \tag{15}$$

The corresponding equations for general rotations about all three coordinate axes are not invertible, *i.e.*, we cannot find

the three rotation angles from (α, α'). Various sets of 3D rotation angles yield the same matching line angles (α, α'). This can be illustrated by two intersecting planes within the 3D object coordinate system that both run through the origin. The first plane represents the $z = 0$ plane, the orientation of the second plane determines the matching line angles (α, α'). Rotating the second plane about the axis that is defined by the matching line of the first plane will not change (α, α'). However, in analogy to the translational components x_0 and y_0 where we were able to gain an equation describing their relationship, the angle pair (α, α') constrains the solution space for the 3D rotation angles.

3. AN ALGORITHM FOR ESTIMATING TRANSFORMATION PARAMETERS

We propose a method that is based on a Fourier transformation to detect the matching lines in the spectra under the integral projection assumption. The method performs a complex division of the frequency values along a line in the spectrum of the first image (projection of the original object) and a second line in the spectrum of the second image (projection of the transformed object). As can be derived from (10), this division yields for matching lines

$$M_2(\xi', \eta') = \frac{F_3(\xi, \eta, 0)}{F_3'(\xi', \eta', 0)} = \frac{F_2(\xi, \eta)}{F_2'(\xi', \eta')} = e^{-j(\xi' x_0 + \eta' y_0)}. \tag{16}$$

The values of $F_2(\xi, \eta)$ and $F_2'(\xi', \eta')$ along the matching lines can be expressed by one-dimensional functions $F_1(\rho)$ and $F_1'(\rho)$ respectively. The substitutions

$$\xi = \rho \cos(\alpha), \quad \eta = \rho \sin(\alpha) \tag{17}$$

$$\xi' = \rho \cos(\alpha'), \quad \eta' = \rho \sin(\alpha') \tag{18}$$

describe the relationship of the two-dimensional frequencies (ξ, η), (ξ', η') and the one-dimensional frequency ρ along matching lines. The variables α, α' were introduced in Section 2.4 and denote the angle of the matching lines with respect to the ξ, ξ'-axes of the $\xi\eta$, $\xi'\eta'$-frequency coordinate systems. Using these substitutions, the spectra $F_2(\xi, \eta)$, $F_2'(\xi', \eta')$ and therefore $M_2(\xi', \eta')$ can now be transformed into one-dimensional representations $F_1(\rho)$, $F_1'(\rho)$ and $M_1(\rho)$ which finally results in

$$\begin{aligned} M_1(\rho) &= M_2(\rho \cos(\alpha'), \rho \sin(\alpha')) \\ &= e^{-j\rho(x_0 \cos(\alpha') + y_0 \sin(\alpha'))} \end{aligned} \tag{19}$$

As can be easily seen from (19), the inverse Fourier transform of $M_1(\rho)$ along matching lines yields a delta function.

$$\begin{aligned} m_1(k) &= \mathcal{F}^{-1}\{M_1(\rho)\} \\ &= \delta(k - \cos(\alpha')x_0 - \sin(\alpha')y_0) \end{aligned} \tag{20}$$

The matching technique matches lines based on the presence of this delta impulse. Equation (20) shows that the location of the delta impulse is determined by the translational components x_0 and y_0. We compute $m_1(k)$ for different values of (α, α') and our objective function is $\max(m_1(k))$.

4. EXPERIMENTAL RESULTS

The analysis of the 3D motion parameters was implemented in MATLAB™ [10]. For generating image sequences, we implemented an application in Visual C++ that is based on OpenGL [11]. Our 3D test object was a flat surface area of size 512×512 originally centered in the xy-plane of the object coordinate system. As a texture, we mapped the well-known *Lena* image onto this surface area. We used no lighting to avoid possible artifacts during the rotation. We projected the object into the 2D image coordinate system using the orthographic projection type in OpenGL (Figure 4(a)). We then rotated the surface area about the origin of the xyz-object coordinate system by specifying various angles for both the azimuth ϕ and the elevation θ and applied the same projection as before. Figures 4(b) and 4(c) show the resulting images for a 3D rotation using $\phi = 60\,^\circ$, $\theta = 0\,^\circ$ and $\phi = 45\,^\circ$, $\theta = 20\,^\circ$ respectively. As can be easily seen from Figures 4(c) and 4(a), images that correspond to the rotated orientation of the 3D object are generally sheared versions of the images that correspond to the 3D object in its original orientation.

(a) $[0\,^\circ, 0\,^\circ]$ (b) $[60\,^\circ, 0\,^\circ]$ (c) $[45\,^\circ, 20\,^\circ]$

Fig. 4. Projections of Lena planes with various 3D orientations

In the matching algorithm, the sampling resolution along the lines was one horizontal/vertical pixel distance, the angular resolution was one degree. We used a bilinear interpolation scheme to determine the complex values of the frequency data along the appropriate lines. We compared the experimental results against the theoretically computed results and found a strong agreement.

Table 1. Angles [in °] of the matching lines for various rotation angles [in °]

ϕ	θ	α_t	α_t'	α_e	α_e'
0	60	0	0	0	0
60	0	90	90	90	90
15	15	44.0070	45.9930	44	46
30	30	40.8934	49.1066	41	49
45	20	62.7637	71.1183	63	72

Table 1 shows the theoretical values α_t, α'_t and the experimental results α_e, α'_e respectively. If either the azimuth or the elevation angle was zero, the exact results were received. Deviations between theoretical and test results are caused by windowing effects. Table 1 shows that these deviations are small even for large rotation angles ϕ and θ.

Figures 5(A) and (B) show the peak distribution for the rotation angles $\phi = 45\,^\circ$ and $\theta = 20\,^\circ$ as a result of the IFFT-Detection method. Figure 5(B) plots the peak distributions inside the rectangular patch of Figure 5(A). The axes represent the angles α and α' from 0° to 360° for the two images. The maximum peak can be clearly isolated. The large components in the neighbourhood of the maximum peak are due to correlation in the frequency data.

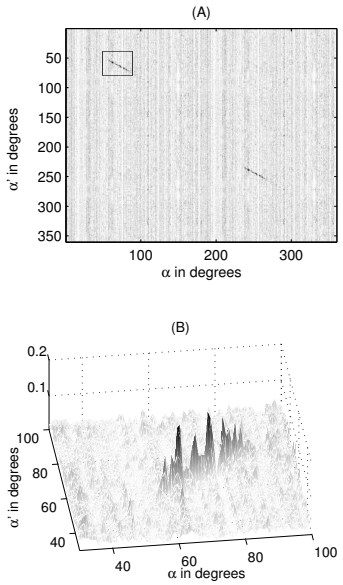

Fig. 5. Evaluation of matching functions

5. CONCLUSION AND FUTURE WORK

In this research, we proposed an integral projection model for the analysis of 3D rigid body transformations. As we showed, there is a strong correspondence between the integral projection model and parallel projection. We developed an algorithm that extracts information about the object transformation from two projected images of the object. The algorithm evaluates a matching function to detect matching lines in the image spectra.

Initial tests were based on the 3D rotation of a textured plane in the 3D object space about various angles. The results of these tests corroborated the theoretical model very well. The proposed matching technique proved to be very suitable for the extraction of 3D transformation information.

Our future research in this area will be focused on optimising both the accuracy and the speed of the detection methods. One of the main research challenges will be to exploit additional structure in the frequency domain in order to extract more information on the 3D rigid body transformation.

6. ACKNOWLEDGEMENTS

The authors would like to thank Andrew Bradley and John Williams (School of ITEE, University of Queensland, Australia) for the helpful discussions in this research.

7. REFERENCES

[1] E. Trucco and A. Verri, *Introductory Techniques for 3D Computer Vision*. Prentice Hall, 1998.

[2] Q.-T. Luong and O. Faugeras, "Camera calibration, scene motion and structure recovery from point correspondences and fundamental matrices," *The International Journal of Computer Vision*, pp. 261–289, 1997.

[3] J. Oliensis and Y. Genc, "New algorithms for two-frame structure from motion," *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 2, pp. 737–744, Sept. 1999.

[4] G. Qian and R. Chellappa, "Structure from motion using sequential monte carlo methods," *Proceedings of the Eigth IEEE International Conference on Computer Vision*, vol. 2, pp. 614–621, July 2001.

[5] C. Leung and B. C. Lovell, "3d reconstruction through segmentation of multi-view image sequences," *APRS Workshop on Digital Image Computing*, pp. 87–92, 2003.

[6] O. Faugeras, *Three-Dimensional Computer Vision*. The MIT Press, 1993.

[7] B. S. Reddy and B. N. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," *IEEE Transactions on Image Processing*, vol. 5(8), pp. 1266–1271, Aug. 1996.

[8] R. N. Bracewell, *Two-Dimensional Imaging*. Prentice-Hall, 1995.

[9] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*. Society of Industrial and Applied Mathematics, 2001.

[10] Mathworks, *http://www.mathworks.com*.

[11] *OpenGL Programming Guide*. Addison-Wesley, 2002.