Projective Alignment of Range and Parallax Data

Miles Hansard, Radu Horaud, Michel Amat, Seungkyu Lee

Projective Alignment of Range and Parallax Data

Miles Hansard[†] Radu Horaud[†] Michel Amat[†] Seungkyu Lee[‡]

miles.hansard@inrialpes.fr
[†] INRIA Grenoble Rhône-Alpes,
655 Avenue de l'Europe,
38330 Montbonnot, France.

[‡] 3D Mixed Reality Group,
Advanced Media Laboratory,
Samsung Advanced Institute of Technology,
South Korea.

Abstract

An approximately Euclidean representation of the visible scene can be obtained directly from a range, or 'time-of-flight', camera. An uncalibrated binocular system, in contrast, gives only a projective reconstruction of the scene. This paper analyzes the geometric mapping between the two representations, without requiring an intermediate calibration of the binocular system. The mapping can be found by either of two new methods, one of which requires point-correspondences between the range and colour cameras, and one of which does not. It is shown that these methods can be used to reproject the range data into the binocular images, which makes it possible to associate high-resolution colour and texture with each point in the Euclidean representation.

1. Introduction

The 3-D structure of a scene can be reconstructed from two or more views, via the *parallax* between corresponding image points. Alternatively, a *range* or 'time of flight' sensor can be used to measure the 3-D structure directly. These two approaches have quite different properties.

The parallax data are hard to obtain, owing to the difficulty of putting the image points into correspondence. Indeed, it may be impossible to find any correspondences in untextured regions. Furthermore, if a Euclidean reconstruction is required, then the cameras must be calibrated. The accuracy of the resulting reconstruction will also tend to decrease with the distance of the scene from the cameras [21].

The range data, on the other hand, are often very noisy (and, for very scattering surfaces, incomplete). The spatial resolution of current range sensors is relatively low, the depth-range is limited, and the luminance signal may be unusable for rendering. It should also be noted that range sensors of the type used here [17] cannot be used in outdoor lighting conditions.

These considerations show that it would be advantageous to combine the range and parallax approaches, in a mixed system [16]. Such a system could, in principle, be used to make high-resolution Euclidean reconstructions, with full photometric information [15]. The task of camera calibration would be simplified by the range sensors, while the visual quality of the reconstruction would be ensured by the colour cameras.

In order to make full use of a mixed range/parallax system, it is necessary to find the exact geometric relationship between the different devices. In particular, the reprojection of the range data, into the colour images, must be obtained. This paper is concerned with the estimation of these geometric relationships. Specifically, the aim is to align the range and parallax reconstructions, by a suitable 3-D transformation. The alignment problem has been addressed previously, by fully calibrating the binocular system, and then aligning the two reconstructions by a rigid transformation [11, 24, 25, 7]. This approach can be extended in two ways. Firstly, it is possible to optimize over an explicit parameterization of the camera matrices, as in the work of Beder et al. [3] and Koch et al. [14]. The relative position and orientation of all cameras can be estimated by this method. Secondly, it is possible to minimize an intensity cost between the images and the luminance signal of the range camera. This method estimates the photometric, as well as geometric, relationships between the different cameras [12, 19, 22]. A complete calibration method, which incorporates all of these considerations, is described by Lindner et al. [16].

The approaches described above, while capable of producing good results, have some limitations. Firstly, there may be residual distortions in the range data, that make a rigid alignment impossible [13]. Secondly, these approaches require the binocular system to be fully calibrated, and re-calibrated after any movement of the cameras. This requires, for best results, many views of a known calibration object. Typical view-synthesis applications, in contrast, re-

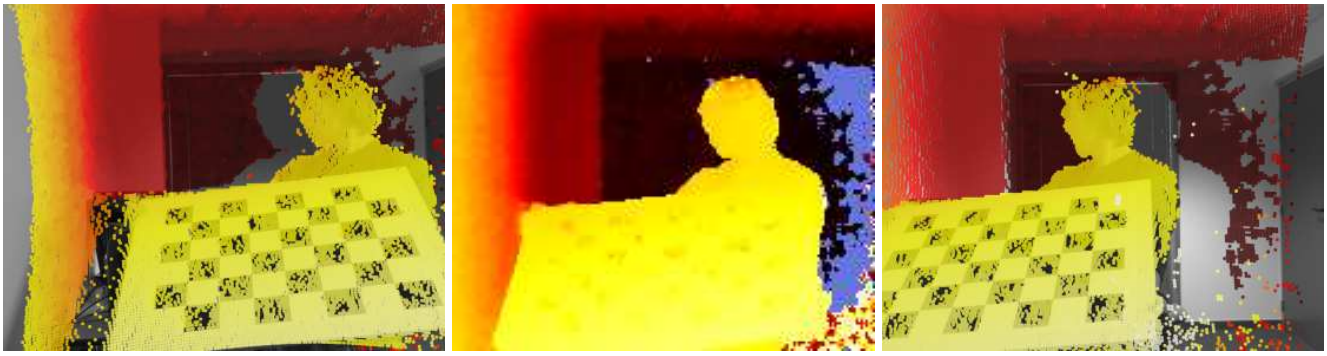

Figure 1. The central panel shows a range image, colour-coded according to depth (the blue region is beyond the far-limit of the device). The left and right cameras were aligned to the range system, using the methods described here. Each 3-D range-pixel is reprojected into the high-resolution left and right images (untinted regions were occluded, or otherwise missing, from the range images). Note the large difference between the binocular views, which would be problematic for dense stereo-matching algorithms. It can also be seen that the range information is noisy, and of low resolution.

quire only a weak calibration of the cameras. One way to remove the calibration requirement is to perform an essentially 2-D registration of the different images [1, 4]. This, however, can only provide an instantaneous solution, because changes in the scene-structure produce corresponding changes in the image-to-image mapping.

An alternative approach is proposed here. It is hypothesized that the range reconstruction is approximately Euclidean. This means that an *uncalibrated* binocular reconstruction can be mapped directly into the Euclidean frame, by a suitable 3-D projective transformation. This is a great advantage for many applications, because automatic uncalibrated reconstruction is relatively easy. Furthermore, although the projective model is much more general than the rigid model, it preserves many important relationships between the images and the scene (e.g. epipolar geometry and incidence of points on planes). Finally, if required, the projective alignment can be upgraded to a fully calibrated solution, as in the methods described above.

It is emphasized that the goal of this work is *not* to achieve the best possible photogrammetric reconstruction of the scene. Rather, the goal is to develop a practical way to associate colour and texture information to each range point, as in figure 1. This output is intended for use in view-synthesis applications.

1.1. Overview and Contributions

The paper is organized as follows. Section 2.1 briefly reviews some standard material on projective reconstruction, while section 2.2 describes the representation of range data in the present work. The chief contributions of the subsequent sections are as follows:

Section 2.3: A *point-based* method that maps an ordinary *projective* reconstruction of the scene onto the corresponding range representation. This does not require the colour cameras to be calibrated (although it may be necessary to correct for lens distortion). Any planar object can be used to find the alignment, provided that image-features can be matched across all views (including that of the range camera).

Section 2.4: A dual *plane-based* method, which performs the same projective alignment, but that does not require any point-matches between the views. Any planar object can be used, provided that it has a simple polygonal boundary that can be segmented in the colour and range data. This is a great advantage, owing to the very low resolution of the luminance data provided by the range camera (176×144 here). It is difficult to automatically extract and match point-descriptors from these images. Furthermore, there are range devices that do not provide a luminance signal at all.

The Euclidean accuracy of the range reconstruction is evaluated in section 3.1, by comparing it to a fully calibrated binocular reconstruction. The new projective methods are evaluated in section 3.2. Conclusions and future directions are discussed in section 4.

2. Methods

This section describes the theory of projective alignment, using the following notation. Bold type will be used for vectors and matrices. In particular, points $\boldsymbol{P}$, $\boldsymbol{Q}$ and planes $\boldsymbol{U}, \boldsymbol{V}$ in the 3-D scene will be represented by column-vectors of homogeneous coordinates, e.g.

$$\boldsymbol{P} = \begin{pmatrix} \boldsymbol{P}_\triangle \\ P_4 \end{pmatrix} \quad \text{and} \quad \boldsymbol{U} = \begin{pmatrix} \boldsymbol{U}_\triangle \\ U_4 \end{pmatrix} \tag{1}$$

where $\boldsymbol{P}_\triangle = (P_1, P_2, P_3)^\top$ and $\boldsymbol{U}_\triangle = (U_1, U_2, U_3)^\top$. The homogeneous coordinates are defined up to a nonzero scaling; for example, $\boldsymbol{P} \simeq (\boldsymbol{P}_\triangle / P_4, 1)^\top$. In particular, if

$P_4 = 1$, then $\boldsymbol{P}_\triangle$ contains the ordinary space coordinates of the point $\boldsymbol{P}$. Furthermore, if $|\boldsymbol{U}_\triangle| = 1$, then U_4 is the signed perpendicular distance of the plane $\boldsymbol{U}$ from the origin, and $\boldsymbol{U}_\triangle$ is the unit normal. The point $\boldsymbol{P}$ is on the plane $\boldsymbol{U}$ if $\boldsymbol{U}^\top \boldsymbol{P} = 0$. The cross product $\boldsymbol{u} \times \boldsymbol{v}$ is often expressed as $(\boldsymbol{u})_\times \boldsymbol{v}$, where $(\boldsymbol{u})_\times$ is a 3×3 antisymmetric matrix. The column-vector of N zeros is written $\boldsymbol{0}_N$.

Projective cameras are represented by 3×4 matrices. For example, the range projection is

$$\boldsymbol{q} \simeq \boldsymbol{C}\boldsymbol{Q} \quad \text{where} \quad \boldsymbol{C} = \big(\boldsymbol{A}_{3\times 3} \,|\, \boldsymbol{b}_{3\times 1}\big). \tag{2}$$

The left and right colour cameras $\boldsymbol{C}_\ell$ and $\boldsymbol{C}_r$ are similarly defined, e.g. $\boldsymbol{p}_\ell \simeq \boldsymbol{C}_\ell \boldsymbol{P}$. Table 1 summarizes the geometric objects that will be aligned.

	Observed Points	Reconstructed Points	Reconstructed Planes
Binocular $\boldsymbol{C}_\ell, \boldsymbol{C}_r$	$\boldsymbol{p}_\ell$, $\boldsymbol{p}_r$	$\boldsymbol{P}$	$\boldsymbol{U}$
Range $\boldsymbol{C}$	$(\boldsymbol{q}, \rho)$	$\boldsymbol{Q}$	$\boldsymbol{V}$

Table 1. Summary of notations in the left, right and range systems.

Points and planes in the two systems are related by the unknown 4×4 space-homography $\boldsymbol{H}$, so that

$$\boldsymbol{Q} \simeq \boldsymbol{H}\boldsymbol{P} \quad \text{and} \quad \boldsymbol{V} \simeq \boldsymbol{H}^{-\top}\boldsymbol{U}. \tag{3}$$

This model encompasses all rigid, similarity and affine transformations in 3-D. It preserves *collinearity* and *flatness*, and is linear in homogeneous coordinates. Note that, in the reprojection process, $\boldsymbol{H}$ can be interpreted as a modification of the camera matrices, e.g. $\boldsymbol{p}_\ell \simeq \big(\boldsymbol{C}_\ell \boldsymbol{H}^{-1}\big)\boldsymbol{Q}$, where $\boldsymbol{H}^{-1}\boldsymbol{Q} \simeq \boldsymbol{P}$.

2.1. Projective Reconstruction

A projective reconstruction of the scene can be obtained from matched points $\boldsymbol{p}_{\ell k}$ and $\boldsymbol{p}_{rk}$, together with the fundamental matrix $\boldsymbol{F}$, where $\boldsymbol{p}_{rk}^\top \boldsymbol{F}\, \boldsymbol{p}_{\ell k} = 0$. The fundamental matrix can be estimated automatically, using the well-established RANSAC method. The camera matrices can then be determined, up to a four-parameter projective ambiguity [10]. In particular, from $\boldsymbol{F}$ and the epipole $\boldsymbol{e}_r$, the cameras can be defined as

$$\boldsymbol{C}_\ell^0 \simeq \big(\boldsymbol{I} \,|\, \boldsymbol{0}_3\big) \;\text{ and }\; \boldsymbol{C}_r^0 \simeq \big((\boldsymbol{e}_r)_\times \boldsymbol{F} + \boldsymbol{e}_r \boldsymbol{g}^\top \,\big|\, \gamma \boldsymbol{e}_r\big). \tag{4}$$

where $\gamma \neq 0$ and $\boldsymbol{g} = (g_1,\, g_2,\, g_3)^\top$ can be used to bring the cameras into a plausible form. This makes it easier to visualize the projective reconstruction and, more importantly, can improve the numerical conditioning of subsequent procedures.

2.2. Range Fitting

The range-camera $\boldsymbol{C}$ provides the distance ρ of each scene-point from the camera-centre, as well as its image-coordinates $\boldsymbol{q} = (x, y, 1)$. The back-projection of this point into the scene is

$$\boldsymbol{Q}_\triangle = \boldsymbol{A}^{-1}\big((\rho/\alpha)\, \boldsymbol{q} - \boldsymbol{b}\big) \quad \text{where} \quad \alpha = \big|\boldsymbol{A}^{-1}\, \boldsymbol{q}\big|. \tag{5}$$

Hence the point $(\boldsymbol{Q}_\triangle, 1)^\top$ is at distance ρ from the optical centre $-\boldsymbol{A}^{-1}\boldsymbol{b}$, in the direction $\boldsymbol{A}^{-1}\boldsymbol{q}$. The scalar α serves to normalize the direction-vector. This is the standard pinhole model, as used in [2].

The range data are noisy and incomplete, owing to illumination and scattering effects. This means that, given a sparse set of features in the intensity image (of the range device), it is not advisable to use the back-projected point (5) directly. A better approach is to segment the image of the plane in each range device (using the the range and/or intensity data). It is then possible to back-project *all* of the enclosed points, and to robustly fit a plane $\boldsymbol{V}_j$ to the enclosed points $\boldsymbol{Q}_{ij}$, so that $\boldsymbol{V}_j^\top \boldsymbol{Q}_{ij} \approx 0$ if point i lies on plane j. Now, the back-projection $\boldsymbol{Q}_\star$ of each sparse feature point $\boldsymbol{q}$ can be obtained by intersecting the corresponding ray with the plane $\boldsymbol{V}$, so that the new range estimate $\rho_\star$ is

$$\rho_\star = \frac{\boldsymbol{V}_\triangle^\top \boldsymbol{A}^{-1}\boldsymbol{b} - V_4}{(1/\alpha)\; \boldsymbol{V}_\triangle^\top \boldsymbol{A}^{-1}\boldsymbol{q}} \tag{6}$$

where $|V_4|$ is the distance of the plane to the camera centre, and $\boldsymbol{V}_\triangle$ is the unit-normal of the range plane. The new point $\boldsymbol{Q}_\star$ is obtained by substituting $\rho_\star$ into (5).

2.3. Point-Based Alignment

It is straightforward to show that the transformation $\boldsymbol{H}$ in (3) could be estimated from five binocular points $\boldsymbol{P}_k$, together with the corresponding range points $\boldsymbol{Q}_k$. This would provide 5×3 equations, which determine the 4×4 entries of $\boldsymbol{H}$, subject to an overall projective scaling. It is better, however, to use the 'Direct Linear Transformation' method [10], which fits $\boldsymbol{H}$ to *all* of the data. This method is based on the fact that if

$$\boldsymbol{P}' = \boldsymbol{H}\boldsymbol{P} \tag{7}$$

is a perfect match for $\boldsymbol{Q}$, then $\mu \boldsymbol{Q} = \lambda \boldsymbol{P}'$, and the scalars λ and μ can be eliminated between pairs of the four implied equations [6]. This results in $\binom{4}{2} = 6$ interdependent constraints per point. It is convenient to write these homogeneous equations as

$$\begin{pmatrix} Q_4 \boldsymbol{P}'_\triangle - P'_4 \boldsymbol{Q}_\triangle \\ \boldsymbol{Q}_\triangle \times \boldsymbol{P}'_\triangle \end{pmatrix} = \boldsymbol{0}_6. \tag{8}$$

Note that if $\boldsymbol{P}'$ and $\boldsymbol{Q}$ are normalized so that $P'_4 = 1$ and $Q_4 = 1$, then the magnitude of the top half of (8) is simply

the distance between the points. Following Förstner [8], the left-hand side of (8) can be expressed as $\big(\boldsymbol{Q}\big)_{\wedge}\boldsymbol{P}'$ where

$$\big(\boldsymbol{Q}\big)_{\wedge} = \begin{pmatrix} Q_4\boldsymbol{I}_3 & -\boldsymbol{Q}_{\triangle} \\ \big(\boldsymbol{Q}_{\triangle}\big)_{\times} & \boldsymbol{0}_3 \end{pmatrix} \tag{9}$$

is a 6×6 matrix, and $\big(\boldsymbol{Q}_{\triangle}\big)_{\times}\boldsymbol{P}_{\triangle} = \boldsymbol{Q}_{\triangle}\times\boldsymbol{P}_{\triangle}$, as usual. The equations (8) can now be written in terms of (7) and (9) as

$$\big(\boldsymbol{Q}\big)_{\wedge}\boldsymbol{H}\boldsymbol{P} = \boldsymbol{0}_6. \tag{10}$$

This system of equations is linear in the unknown entries of $\boldsymbol{H}$, the columns of which can be stacked into the 16×1 vector $\boldsymbol{h}$. The Kronecker product identity $\mathrm{vec}(\boldsymbol{XYZ}) = (\boldsymbol{Z}^{\top}\otimes\boldsymbol{X})\,\mathrm{vec}(\boldsymbol{Y})$ can now be applied, to give

$$\Big(\boldsymbol{P}^{\top}\otimes\big(\boldsymbol{Q}\big)_{\wedge}\Big)\boldsymbol{h} = \boldsymbol{0}_6 \quad \text{where} \quad \boldsymbol{h} = \mathrm{vec}\big(\boldsymbol{H}\big). \tag{11}$$

If M points are observed on each of N planes, then there are $k = 1,\ldots,MN$ observed pairs of points, $\boldsymbol{P}_k$ from the projective reconstruction and $\boldsymbol{Q}_k$ from the range back-projection. The MN corresponding 6×16 matrices $\big(\boldsymbol{P}_k^{\top}\otimes(\boldsymbol{Q}_k)_{\wedge}\big)$ are stacked together, to give the complete system

$$\begin{pmatrix} \boldsymbol{P}_1^{\top} \;\otimes\; \big(\boldsymbol{Q}_1\big)_{\wedge} \\ \vdots \\ \boldsymbol{P}_{MN}^{\top}\otimes\big(\boldsymbol{Q}_{MN}\big)_{\wedge} \end{pmatrix}\boldsymbol{h} = \boldsymbol{0}_{6MN} \tag{12}$$

subject to the constraint $|\boldsymbol{h}| = 1$, which excludes the trivial solution $\boldsymbol{h} = \boldsymbol{0}_{16}$. It is straightforward to obtain an estimate of $\boldsymbol{h}$ from the SVD of the the $6MN\times 16$ matrix on the left of (12). This solution, which minimizes an *algebraic error* [10], is the singular vector corresponding to the smallest singular value of the matrix. In the minimal case, $M = 1, N = 5$, the matrix would be 30×16. Note that the point coordinates should be transformed, to ensure that (12) is numerically well-conditioned [10]. In this case the transformation ensures that $\sum_k \boldsymbol{P}_{k\triangle} = \boldsymbol{0}_3$ and $\frac{1}{MN}\sum_k |\boldsymbol{P}_{k\triangle}| = \sqrt{3}$, where $P_{k4} = 1$. The analogous transformation is applied to the range points $\boldsymbol{Q}_k$.

The DLT method, as will be shown in section 3.2, gives a reasonable approximation $\boldsymbol{H}_0$ of the homography (3). This can be used as a starting-point for the iterative minimization of a more appropriate error measure. In particular, consider the *reprojection error* in the left image,

$$E_{\ell}(\boldsymbol{C}_{\ell}) = \sum_{k=1}^{MN} D\big(\boldsymbol{C}_{\ell}\boldsymbol{Q}_k,\, \boldsymbol{p}_{\ell k}\big)^2 \tag{13}$$

where $D(\boldsymbol{p},\boldsymbol{q}) = |\boldsymbol{p}_{\triangle}/p_3 - \boldsymbol{q}_{\triangle}/q_3|$. A 12-parameter optimization of (13), starting with $\boldsymbol{C}_{\ell}^1 \leftarrow \boldsymbol{C}_{\ell}^0\boldsymbol{H}_0^{-1}$, can be performed by the Levenberg-Marquardt algorithm [18]. The result will be the camera matrix $\boldsymbol{C}_{\ell}^{\star}$ that best reprojects the range data into the left image ($\boldsymbol{C}_r^{\star}$ is similarly obtained). The solution, provided that the calibration points adequately covered the scene volume, will remain valid for subsequent depth and range data.

Alternatively, it is possible to minimize the *joint* reprojection error,

$$E\big(\boldsymbol{H}^{-1}\big) = E_{\ell}\big(\boldsymbol{C}_{\ell}^0\boldsymbol{H}^{-1}\big) + E_r\big(\boldsymbol{C}_r^0\boldsymbol{H}^{-1}\big) \tag{14}$$

over the (inverse) homography $\boldsymbol{H}^{-1}$. The 16 parameters are again minimized by the Levenberg-Marquardt algorithm, starting from the DLT solution $\boldsymbol{H}_0^{-1}$.

The difference between the separate (13) and joint (14) minimizations is that the latter preserves the original epipolar geometry, whereas the former does not. Recall that $\boldsymbol{C}_{\ell}$ $\boldsymbol{C}_r$, $\boldsymbol{H}$ and $\boldsymbol{F}$ are all defined up to scale, and that $\boldsymbol{F}$ satisfies an additional rank-two constraint [10]. Hence the underlying parameters can be counted as $(12-1)+(12-1) = 22$ in the separate minimizations, and as $(16-1) = 15$ in the joint minimization. The fixed epipolar geometry accounts for the $(9-2)$ missing parameters in the joint minimization. If $\boldsymbol{F}$ is known to be very accurate (or must be preserved) then the joint minimization (14) should be performed. This will also preserve the original binocular triangulation, provided that a projective-invariant method was used [9]. However, if minimal reprojection error is the objective, then the cameras should be treated separately. This will lead to a new fundamental matrix $\boldsymbol{F}^{\star} = (\boldsymbol{e}_r^{\star})_{\times}\boldsymbol{C}_r^{\star}(\boldsymbol{C}_{\ell}^{\star})^{+}$, where $(\boldsymbol{C}_{\ell}^{\star})^{+}$ is the generalized inverse. The right epipole is obtained from $\boldsymbol{e}_r^{\star} = \boldsymbol{C}_r^{\star}\boldsymbol{d}_{\ell}^{\star}$, where $\boldsymbol{d}_{\ell}^{\star}$ represents the nullspace $\boldsymbol{C}_{\ell}^{\star}\boldsymbol{d}_{\ell}^{\star} = \boldsymbol{0}_3$.

2.4. Plane-Based Alignment

The DLT algorithm of section 2.3 can also be used to recover $\boldsymbol{H}$ from matched *planes*, rather than matched points. Equation (10) becomes

$$\big(\boldsymbol{V}\big)_{\wedge}\,\boldsymbol{H}^{-\top}\boldsymbol{U} = \boldsymbol{0}_6 \tag{15}$$

where $\boldsymbol{U}$ and $\boldsymbol{V}$ represent the estimated coordinates of the same plane in the parallax and range reconstructions, respectively. The estimation procedure is identical to that in section 2.3, but with $\mathrm{vec}(\boldsymbol{H}^{-\top})$ as the vector of unknowns.

This method, in practice, produces very poor results. The chief reason that obliquely-viewed planes are foreshortened, and therefore hard to detect/estimate, in the low-resolution range images. It follows that the calibration dataset is biased towards fronto-parallel planes.[1] This bias allows the registration to slip sideways, perpendicular to the primary direction of the range camera. The situation is greatly improved by assuming that the *boundaries* of the

[1] The point-based algorithm is unaffected by this bias, because the scene is ultimately 'filled' with points, regardless of the contributing planes.

planes can be detected. For example, if the calibration object is rectangular, then the range-projection of the plane $\boldsymbol{V}$ is bounded by four edges $\bar{\boldsymbol{v}}_i$, where $i = 1, \dots 4$. Note that these are detected as *depth* edges, and so no luminance data are required. The edges, represented as lines $\bar{\boldsymbol{v}}_i$, back-project as the faces of a pyramid,

$$\overline{\boldsymbol{V}}_i = \boldsymbol{C}^\top \bar{\boldsymbol{v}}_i = \begin{pmatrix} \overline{\boldsymbol{V}}_{i\triangle} \\ 0 \end{pmatrix}, \quad i = 1, \dots L \tag{16}$$

where $L = 4$ in the case of a quadrilateral projection. These planes are linearly dependent, because they pass through the centre of projection; hence the fourth coordinates are all zero if, as here, the range camera is at the origin. Next, if the corresponding edges $\bar{\boldsymbol{u}}_{\ell i}$ and $\bar{\boldsymbol{u}}_{ri}$ can be detected in the binocular system, using both colour and parallax information, then the planes $\overline{\boldsymbol{U}}_i$ can easily be constructed. Each calibration plane now contributes an additional $6L$ equations

$$\big(\overline{\boldsymbol{V}}_i\big)_\wedge \, \boldsymbol{H}^{-\top} \overline{\boldsymbol{U}}_i = \boldsymbol{0}_6 \tag{17}$$

to the DLT system (12). Although these equations are quite redundant (any two planes span all possibilities), they lead to a much better DLT estimate. This is because they represent exactly those planes that are most likely to be missed in the calibration data, owing to the difficulty of feature-detection over surfaces that are extremely foreshortened in the image.

As in the point-based method, the plane coordinates should be suitably transformed, in order to make the numerical system (12) well-conditioned. The transformed coordinates satisfy the location constraint $\sum_k \boldsymbol{U}_{k\triangle} = \boldsymbol{0}_3$, as well as the scale constraint $\sum_k |\boldsymbol{U}_{k\triangle}|^2 = 3 \sum_k U_{k4}^2$, where $\boldsymbol{U}_{k\triangle} = (U_{k1}, U_{k2}, U_{k3})^\top$, as usual. A final renormalization $|\boldsymbol{U}_k| = 1$ is also performed. This procedure, which is also applied to the $\boldsymbol{V}_k$, is analogous to the treatment of line-coordinates in DLT methods [23].

The remaining problem is that the original reprojection error (13) cannot be used to optimize the solution, because no luminance features $\boldsymbol{q}$ have been detected in the range images (and so no 3-D points $\boldsymbol{Q}$ have been distinguished). This can be solved by reprojecting the physical edges of the calibration planes, after reconstructing them as follows. Each edge-plane $\overline{\boldsymbol{V}}_i$ intersects the range plane $\boldsymbol{V}$ in a space-line, represented by the 4×4 Plücker matrix

$$\boldsymbol{W}_i = \boldsymbol{V}\, \overline{\boldsymbol{V}}_i^\top - \overline{\boldsymbol{V}}_i\, \boldsymbol{V}^\top. \tag{18}$$

The line $\boldsymbol{W}_i$ reprojects to a 3×3 antisymmetric matrix [10]; for example $\boldsymbol{W}_{\ell i} \simeq \boldsymbol{C}_\ell\, \boldsymbol{W}_i\, \boldsymbol{C}_\ell^\top$ in the left image, and similarly in the right. Note that $\boldsymbol{W}_{\ell i} \boldsymbol{p}_\ell = \boldsymbol{0}$ if the point $\boldsymbol{p}_\ell$ is on the reprojected line [10]. The line-reprojection error can therefore be written as

$$E_\ell^\times(\boldsymbol{C}_\ell) = \sum_{i=1}^{L} \sum_{j=1}^{N} D_\times\big(\boldsymbol{C}_\ell\, \boldsymbol{W}_i\, \boldsymbol{C}_\ell^\top,\, \bar{\boldsymbol{u}}_{\ell ij}\big)^2. \tag{19}$$

The function $D_\times\big(\boldsymbol{M}, \boldsymbol{n}\big)$ compares image-lines, by computing the sine of the angle between the two coordinate-vectors,

$$D_\times(\boldsymbol{M}, \boldsymbol{n}) = \frac{\sqrt{2}\,|\boldsymbol{M}\boldsymbol{n}|}{|\boldsymbol{M}|\,|\boldsymbol{n}|} = \frac{|\boldsymbol{m} \times \boldsymbol{n}|}{|\boldsymbol{m}|\,|\boldsymbol{n}|}, \tag{20}$$

where $\boldsymbol{M} = (\boldsymbol{m})_\times$, and $|\boldsymbol{M}|$ is the Frobenius norm. It is emphasized that the coordinates *must* be normalized by a suitable transformations $\boldsymbol{G}_\ell$ and $\boldsymbol{G}_r$, as in the case of the DLT. For example, the line $\boldsymbol{n}$ should be fitted to points of the form $\boldsymbol{G}\boldsymbol{p}$, and then $\boldsymbol{M}$ should be transformed as $\boldsymbol{G}^{-\top}\boldsymbol{M}$, before computing (20). The reprojection error (19) is numerically unreliable without this normalization.

The line-reprojection (20) can either be minimized separately for each camera, or jointly as

$$E^\times\big(\boldsymbol{H}^{-1}\big) = E_\ell^\times\big(\boldsymbol{C}_\ell^0 \boldsymbol{H}^{-1}\big) + E_r^\times\big(\boldsymbol{C}_r^0 \boldsymbol{H}^{-1}\big) \tag{21}$$

by analogy with (14). Finally, it should be noted that although (20) is defined in the *image*, it is an *algebraic* error. However, because the errors in question are small, this measure behaves predictably.

3. Experiments

The following experiments are based on images from two colour cameras mounted on a rail (approx. separation 49cm), plus one Mesa Imaging SR4000 range camera [17], mounted approximately midway between. All three optical axes are approximately parallel. A chequerboard calibration object, with 8cm×8cm squares, was used. The basic data set contains 27 image-triples, with 35 vertices put (automatically) into subpixel correspondence across each triple. Lens distortion was corrected, in both colour and range cameras, by standard methods [5]. The task of automatic depth-edge detection is beyond the scope of the present work. Hence the lines used in the reprojection error (19) are formed by joining the four corner-points of each chequerboard.

Section 3.1 analyzes the full data set, whereas the evaluation in 3.2 uses 500 random subsets of the data (recall from section 2 that the projective methods require just five points/planes in the minimal case).

3.1. Similarity Alignment

A complete Euclidean calibration of the binocular system was performed, in order to provide ground-truth for the experiments. The calibrated binocular reconstruction is known to be accurate, and can therefore be used to check for geometric distortion in the range data. The reconstruction was aligned to the range data by a similarity transformation $\boldsymbol{Q} \approx \boldsymbol{S}\boldsymbol{P}$, which is analogous to the uncalibrated case $\boldsymbol{Q} \approx \boldsymbol{H}\boldsymbol{P}$ in (3). The transformation has the form

$$\boldsymbol{S}(\boldsymbol{t}, \theta, \sigma) = \begin{pmatrix} \sigma \boldsymbol{R}(\theta) & \boldsymbol{t} \\ \boldsymbol{0}_{1\times 3} & 1 \end{pmatrix} \tag{22}$$

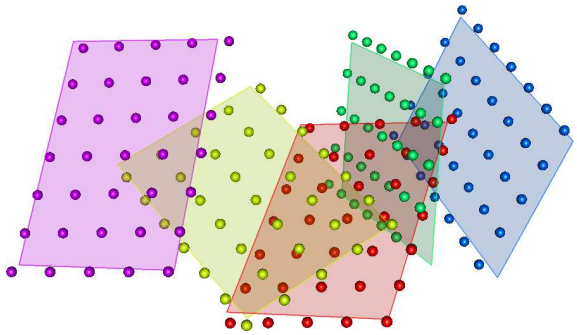

Figure 2. Similarity alignment: All points in the *calibrated* binocular reconstruction were aligned to the range planes, using the procrustes algorithm. The figure shows a random subset of the planes. Each sphere has radius 1cm.

where $\boldsymbol{R}$ is a rotation matrix, $\boldsymbol{t}$ is a translation, and σ is a scale factor. This alignment was performed using the standard procrustes method [20, 11]. A good fit was obtained, as can be seen from the subset of data shown in Fig. 2. The median residual distance was 1.962cm, in a data volume of 110cm^3. The full 3-D alignment shows no significant distortions in the range data, although it is very noisy. This demonstrates that the range reconstruction is nearly Euclidean.

The similarity transformations $\boldsymbol{S}$ are a subgroup of the homographies $\boldsymbol{H}$, and so the new projective methods can also be applied in the calibrated case. The error metrics are, however, different. The procrustes method minimizes the pointwise squared distance, in 3-D. The residual distance is a geometric error, which will be called GE3. By analogy with section 2.3, this solution is used as a starting point for the minimization of the 2-D pointwise reprojection error (GE2, eqn. 13) using the camera matrices $\boldsymbol{C}_\ell^0\boldsymbol{S}^{-1}$ and $\boldsymbol{C}_r^0\boldsymbol{S}^{-1}$. In contrast, the point-based DLT minimizes 3-D algebraic error (AE3,10), followed by GE2. The plane-based DLT minimizes a dual 3-D algebraic error (AE$^\times$3,15), followed by a dual 2-D algebraic error (AE$^\times$2, 19). It is most useful to evaluate each of these solutions on the *geometric* errors, regardless of which error was actually minimized. This information is given in table 2.

Two conclusions can be drawn from the table. Firstly, after optimization of the reprojection error, the projective model gives much better results than the similarity model. Indeed this should be the case, if the optimization performs correctly, because the projective model is more general. Secondly, it can be seen that the initial plane-based DLT solution is by far the worst, owing to the ambiguity of the plane coordinates. Nonetheless, the advantage of the projective model is recovered during the nonlinear minimization, and the final result is almost as good as the point-based projective method. It may also be noted that that GE3 usually increases during the optimization of GE2, except in the case of planes; this shows the weakness of AE$^\times$3.

		GE3 cm	GE2 px	$\lvert\boldsymbol{t}\rvert$ cm	θ°	σ
$\boldsymbol{H}$	AE3	1.323	3.671	27.505	0.676	1.028
	GE2	1.330	2.336	27.202	0.607	1.026
$\boldsymbol{H}^{-\top}$	AE$^\times$3	3.021	26.261	21.400	2.958	1.059
	AE$^\times$2	1.492	2.591	27.411	1.154	1.026
$\boldsymbol{S}$	GE3	2.646	12.614	27.567	0.679	1.027
	GE2	3.117	4.944	26.603	0.932	0.986

Table 2. Residuals and parameters. The three rows summarize the point-based projective, plane-based projective, and point-based similarity methods, respectively. The two sub-rows, in each case, represent the initial linear solution and the subsequent nonlinear solution. The first two columns show the geometric errors of each estimate, followed by the similarity parameters.

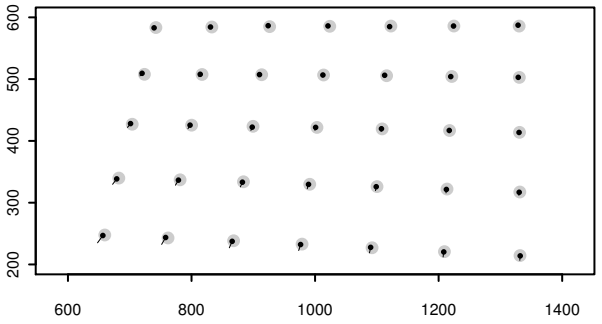

Figure 3. Projective point-based alignment: All planes in the *uncalibrated* binocular reconstruction were aligned to the range data; the initially *worst* case (by pointwise RMS) is shown. The grey discs represent the actual image features. The tails of the dots indicate the initial (DLT) locations of the reprojected range points; the heads indicate the their final locations, after nonlinear optimization over all data.

Table 2 also gives the physical parameters of each estimated transformation (note that the translation is approximately half the binocular baseline, as expected). The homography parameters come from the similarity matrix $\boldsymbol{S}_H$, which is obtained by solving the procrustes problem $\boldsymbol{S}_H\boldsymbol{P}_k \approx \boldsymbol{H}\boldsymbol{P}_k$, over all points $\boldsymbol{P}_k$. Hence $\boldsymbol{S}_H$ is the best similarity-approximation of $\boldsymbol{H}$, over the given data.

3.2. Projective Alignment

It was shown, in the preceding section, that the range data are nearly Euclidean. It was also shown that a projective transformation $\boldsymbol{H}$, which maps a binocular reconstruction to the range data, can be found by the methods of section 2. The following experiments address the practicality of these methods in relation to smaller data sets, *without* binocular calibration.

The first question concerns **reliability**: Given the initial linear solutions, do the iterative algorithms always reach approximately the same level of reprojection error? This was tested by splitting the data into 500 subsets, each containing

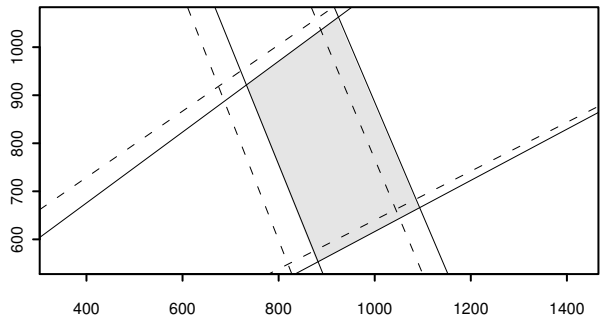

Figure 4. Projective plane-based alignment: All planes in the *uncalibrated* binocular reconstruction were aligned to the range data; the initially *worst* case (by pointwise RMS) is shown. The grey polygon represents the actual image of the plane. The dashed lines represent the initial (DLT) locations of the lines; the solid lines are the final locations, after nonlinear optimization over all data.

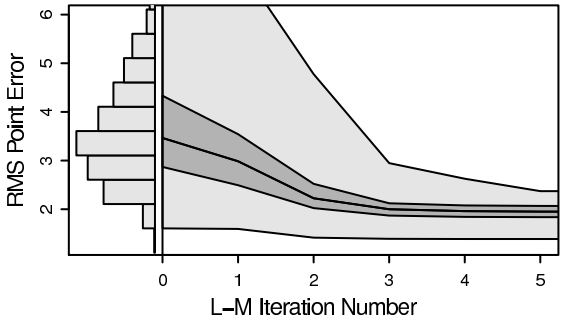

Figure 5. Convergence of the projective point-based method, evaluated on 500 randomly-chosen sets of 10 planes. The histogram shows the distribution of the initial DLT RMSE in the 1224×1624 images. The light polygon is the total envelope of the 500 error traces. The dark polygon is the inter-quartile range, which contains the central 50% of the traces, around the median line.

a random selection of 10 planes. The median reprojection errors, after nonlinear minimization of each camera matrix, were 1.947 pixels (inter-quartile range 0.230) for the point-based method, and 2.302 (IQR 0.280) for the plane based method. Figures 3 and 4 show example fits from both methods.

The convergence of the Levenberg-Marquardt procedure, in all trials, is plotted in Figs. 5 and 6. Note that each plot represents the minimization of 500 *different* cost functions, and so convergence to exactly the same level is not expected. Alternatively, it would be possible to minimize the total reprojection error over *all* data, starting from the 500 initial solutions. Such a test would be artificial because, in reality, all of the available data would be used in both the linear and nonlinear stages of the calibration. Nonetheless, it was verified that all traces converge to the same level in such a test.

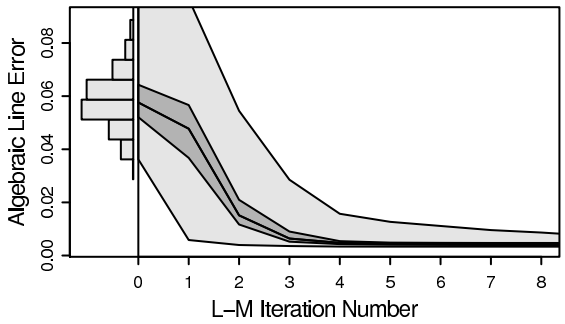

Figure 6. Convergence of the projective plane-based method, evaluated on 500 randomly-chosen sets of 10 planes. Details as in the caption of Fig. 5.

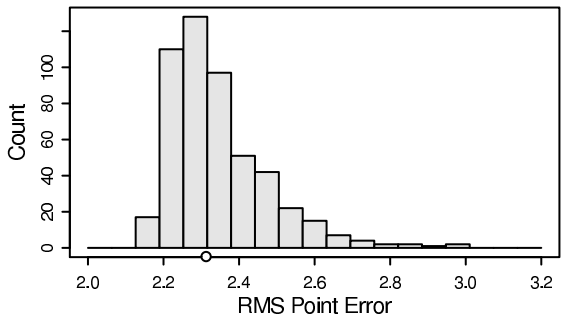

Figure 7. Test-error of the projective point-based method. This shows pointwise RMSE of the 500 optimized solutions, corresponding to the terminations of the lines in Fig. 5. The error is evaluated on the *full* data set, only a small fraction of which was used for each fit. The median value is marked on the horizontal axis.

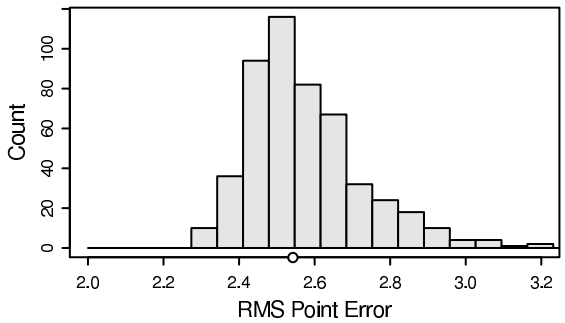

Figure 8. Test-error of the projective plane-based method, corresponding to the terminations of the lines in Fig. 6. Details as in the caption of Fig. 7.

The second question is that of **accuracy**: Do the nonlinear solutions give low reprojection error, even when evaluated on data that were *not* present in the minimization? This was tested by measuring the RMSE of each subset-solution on *all* of the data. Figures 7 and 8 show the results for the point-based and plane-based methods, respectively.

It can be seen that both methods give errors of around 2.5 pixels, in the 1624×1224 images. The point-based method (median error 2.313, IQR 0.158) is slightly more accurate than the plane-based method (median error 2.542, IQR 0.179). This is not surprising, because each plane contributes 35 vertices, but only four edge-lines. Finally, it was verified that subpixel accuracy can be obtained from these solutions, by performing a bundle adjustment over the estimated cameras *and* points. This procedure, however, is not practical if the points arrive at video-rate. The methods described here, which optimize and then fix the cameras, are more useful in such applications.

4. Discussion

It has been shown that there is a projective relationship between the data provided by a range sensor, and an uncalibrated binocular reconstruction. Two practical methods for computing the projective transformation have been introduced; one that requires luminance point-correspondences between the range and colour cameras, and one that does not. Either of these methods can be used to associate binocular colour and texture with each 3-D point in the range reconstruction. The plane-based method, although slightly less accurate, is particularly attractive. This is because, even if the range camera provides a luminance signal, the spatial resolution of current devices is too low for reliable wide-baseline matching.

Future work will address the extension of these methods to larger configurations of several range and colour cameras. The automatic detection of planar calibration objects, especially in the range data, will also be addressed.

References

[1] B. Bartczak and R. Koch. Dense depth maps from low resolution time-of-flight depth and high resolution color views. In *Proc. Int. Symp. on Visual Computing (ISVC)*, pages 228–239, 2009.

[2] C. Beder, B. Bartczak, and R. Koch. A comparison of PMD-cameras and stereo-vision for the task of surface reconstruction using patchlets. In *Proc. CVPR*, pages 1–8, 2007.

[3] C. Beder, I. Schiller, and R. Koch. Photoconsistent relative pose estimation between a PMD 2D3D-camera and multiple intensity cameras. In *Proc. Symp. of the German Association for Pattern Recognition (DAGM)*, pages 264–273, 2008.

[4] A. Bleiweiss and M. Werman. Fusing time-of-flight depth and color for real-time segmentation and tracking. In *Proc. DAGM Workshop on Dynamic 3D Imaging*, pages 58–69, 2009.

[5] G. Bradski. The OpenCV library. *Dr. Dobb's Journal of Software Tools*, 25(11):122–125, 2000.

[6] G. Csurka, D. Demirdjian, and R. Horaud. Finding the collineation between two projective reconstructions. *Computer Vision and Image Understanding*, 75(3):260–268, 1999.

[7] J. M. Dubois and H. Hügli. Fusion of time-of-flight camera point clouds. In *ECCV Workshop on Multi-Camera and Multi-modal Sensor Fusion Algorithms and Applications*, 2008.

[8] W. Förstner. Uncertainty and projective geometry. In E. Bayro-Corrochano, editor, *Handbook of Geometric Computing*, pages 493–534. Springer, 2005.

[9] R. Hartley and P. Sturm. Triangulation. *Computer Vision and Image Understanding*, 68(2):146–157, 1997.

[10] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[11] B. Horn, H. Hilden, and S. Negahdaripour. Closed-form solution of absolute orientation using orthonormal matrices. *J. Optical Society of America A*, 5(7):1127–1135, 1988.

[12] B. Huhle, S. Fleck, and A. Schilling. Integrating 3D time-of-flight camera data and high resolution images for 3DTV applications. In *Proc. 3DTV*, pages 1–4, 2007.

[13] Y. Kim, D. Chan, C. Theobalt, and S. Thrun. Design and calibration of a multi-view TOF sensor fusion system. In *Proc. CVPR Workshop on time-of-flight Camera based Computer Vision*, 2008.

[14] R. Koch, I. Schiller, B. Bartczak, F. Kellner, and K. Köser. MixIn3D: 3D mixed reality with ToF-camera. In *Proc. DAGM workshop on dynamic 3D imaging*, pages 126–141, 2009.

[15] A. Kolb, E. Barth, R. Koch, and R. Larsen. Time-of-flight cameras in computer graphics. *Computer Graphics Forum*, 29(1):141–159, 2010.

[16] M. Lindner, I. Schiller, A. Kolb, and R. Koch. Time-of-flight sensor calibration for accurate range sensing . *Computer Vision and Image Understanding*, 114(12):1318–1328, 2010.

[17] Mesa Imaging AG. `http://www.mesa-imaging.ch`.

[18] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C*. Cambridge University Press, 2nd edition, 1992.

[19] I. Schiller, C. Beder, and R. Koch. Calibration of a PMD camera using a planar calibration object together with a multi-camera setup. In *Int. Arch. Soc. Photogrammetry, Remote Sensing and Spatial Information Sciences XXI*, pages 297–302, 2008.

[20] P. Schönemann and R. Carroll. Fitting one matrix to another under choice of a central dilation and a rigid motion. *Psychometrika*, 35(2):245–255, 1970.

[21] A. Verri and V. Torre. Absolute depth estimate in stereopsis. *J. Optical Society of America A*, 3(3):297–299, 1986.

[22] J. Wu, Y. Zhou, H. Yu, and Z. Zhang. Improved 3D depth image estimation algorithm for visual camera. In *Proc. International Congress on Image and Signal Processing*, 2009.

[23] H. Zeng, X. Deng, and Z. Hu. A new normalized method on line-based homography estimation. *Pattern Recognition Letters*, 29:1236–1244, 2008.

[24] Q. Zhang and R. Pless. Extrinsic calibration of a camera and laser range finder (improves camera calibration). In *Proc. Int. Conf. on Intelligent Robots and Systems*, pages 2301–2306, 2004.

[25] J. Zhu, L. Wang, R. G. Yang, and J. Davis. Fusion of time-of-flight depth and stereo for high accuracy depth maps. In *Proc. CVPR*, pages 1–8, 2008.