

Eliminating False Matches for the Projective Registration of Free-Form Surfaces With Small Translational Motions

Yonghuai Liu

Abstract—In this paper, we make a detailed study of two rigid-motion constraints. The importance of these two constraints is twofold: first, they reveal the inherent relationship between the three-dimensional–two-dimensional (3-D–2-D) point correspondences and the motion parameters of interest; second, they can be used to measure the traditional ICP criterion established point match qualities based on which different point matches can be compared and relatively good point matches can be selected for motion-parameter update in the projective registration of free-form surfaces subject to small translational motions. The experimental results based on both synthetic data and real images have shown that the rigid motion constraints are powerful in evaluating the possible 3-D–2-D point matches established by the traditional ICP criterion, thus achieving encouraging projective registration results.

Index Terms—Colinearity, focus of expansion, free-form surface, ICP, projective registration, rigid motion constraint, rigid pure translational motion, three-dimensional–two-dimensional (3-D–2-D) point correspondence.

I. INTRODUCTION

PROJECTIVE registration of free-form surfaces described as a three-dimensional (3-D) map and a two-dimensional (2-D) projective image taken by a robot at a certain viewpoint is a fundamental problem in the robot vision literature (3-D–2-D registration), as it finds applications in many areas, such as robot location, path planning, obstacle avoidance, object modeling and recognition, and scene understanding [13], [15], [37] etc. The goal of projective registration is twofold: one is to establish 3-D–2-D point correspondences between a 3-D model and its projective image, the other is to estimate camera motion parameters that bring the transformed 3-D model and the projective image into alignment. In practice, these two goals are often entwined, thus complicating projective registration.

A. Related Work

Many methods have been proposed to solve the registration problem based on techniques, such as iterative closest point (ICP) [1], [3], [44], intensity based similarity [32], correlation ratio [33], difference decomposition [12], SSD operator [20], marker matching based interactive method [16], shape matching

[4], feature matching [14], [40], [42], stereo simplex [36], soft-assign [5], [6], and many others. The intrinsic natural and practical idea of the ICP algorithm has attracted much attention from the machine vision and image processing community since it was proposed in 1992 independently by several researchers [1], [3], [44]. Since inaccurate initialization of camera pose, occlusion, appearance and disappearance of points, and noise presence in image data render the traditional ICP criterion (given camera pose rotation matrix \mathbf{R} and translation vector \mathbf{t} , the correspondent of a point \mathbf{p} in the first image is the closest point in the second image to $\mathbf{R}\mathbf{p} + \mathbf{t}$) to inevitably introduce false matches in almost every iteration of registration. As a result, a large number of methods have been proposed either to increase the accuracy of the ICP criterion established point correspondences based on techniques, such as color information [19] and invariant features [35], or to evaluate the ICP criterion established point correspondences based on techniques, such as geometric properties of correspondence vectors [30] and geometric properties of reflected correspondence vectors [34]. These improved methods are mainly used to register two sets of 3-D data (3-D–3-D registration) with or without occlusion, appearance and disappearance of points.

Recently, the traditional ICP algorithm has been extended for use in projective registration. In [24], the edges are first extracted from both reflectance and intensity images. The 3-D–2-D correspondences between the extracted edge points are then established using the traditional ICP algorithm. The camera pose parameters are finally estimated by an M-estimator, combating false matches for more accurate camera pose estimation. In [9], the traditional ICP algorithm has been extended for 3-D–2-D projective registration of free-form curves and surfaces. But the distance is defined as the weighted sum of distance between 2-D curve points and projected 3-D curve points and that between normalized tangent vectors to the 2-D curve and to the projected 3-D curve respectively or as the weighted sum of distance between 2-D surface points and projected 3-D surface points and that between normal vectors to the 2-D surface and to the projected 3-D surface respectively. In order to accelerate the search for closest points, a k-D tree data structure is used. For the evaluation of the established correspondences, the general χ^2 test is then employed. Once the 3-D–2-D point correspondences have been established by the traditional ICP algorithm, the extended Kalman filter is finally used to estimate the camera pose. In [21], [22], a skeleton of a 3-D vessel from a 3-D model and the region of interest in an X-ray image are first extracted and a territory-based method

Manuscript received May 4, 2004; revised October 4, 2004. This work was supported in part by ARPA, in part by NSF, and in part by NASA. This paper was recommended by Associate Editor P. Bhattacharya.

The author is with the Department of Computer Science, The University of Wales, Aberystwyth, Ceredigion SY23 3DB, Wales, U.K. (e-mail: yyl@aber.ac.uk).

Digital Object Identifier 10.1109/TSMCB.2005.843978

is then used to search for the 3-D–2-D point correspondences [23]. The rotation parameters are estimated linearizing the rotation matrix represented with a quaternion and separating the rotation from the translation. In [25], given a good initial estimate of camera pose parameters, the authors propose minimizing the signed distance between the recovered 3-D lines from their projected contours and a closed 3-D surface to refine the camera pose parameters, using the Levenberg-Marquardt method. While the distance between the recovered 3-D lines from their projected contours and the closed 3-D surface is computed as the minimum distance between points on the lines and the surface, its computation is accelerated by the octree spline representation of the distance map. In [43], assuming that all image points can find correspondents in the model, the authors propose minimizing the distances between the transformed 3-D point and the projection rays to establish the 3-D–2-D point correspondences, thus transforming 3-D–2-D motion estimation to 3-D–3-D motion estimation. In order to account for outliers, an M-estimator is used. In [27], the camera pose is estimated by the Newton method through minimizing a cost function of the distance between a point on the projected 3-D skeleton and the intersection point of the 2-D skeleton with the line passing through that point and perpendicular to the tangent of the projected 3-D skeleton. However, the correspondences between 3-D and 2-D skeletons are associated by the user.

All these methods often require human intervention, feature extraction, special object structure, or impose constraints on the image data acquisition. Thus, for automatic registration of free form surfaces, a substantial improvement is still needed.

B. Our Work

Our research has shown that 3-D–2-D projective registration is significantly more difficult than 3-D–3-D registration, since it is more sensitive to the initialization of motion parameters and the motion parameter estimation is more sensitive to both false matches and noise distribution in image points data. To facilitate 3-D–2-D projective registration, the rigid pure translational motion was chosen in this paper. Such camera configuration was also adopted by other algorithms for accurate structural estimation [20], accurate camera pose estimation [26], or accurate camera calibration [17].

Intuitively, the 3-D–2-D point correspondences and the camera pose parameters should be optimized simultaneously for accurate registration results. However, such simultaneous optimization is in practice either difficult to formulate or difficult to implement. Thus, in this paper, we decompose the registration process into three main steps as: (1) Use the traditional ICP criterion to establish possible 3-D–2-D point correspondences; (2) Use rigid motion constraints to evaluate the possible 3-D–2-D point correspondences; and finally (3) Use the Monte Carlo resampling technique to update the camera pose parameters based on the refined 3-D–2-D correspondences and repeat these steps until convergence. Doing so is justified by the following three reasons: (1) the novel approach can be interpreted in the framework of the EM algorithm [7]. In the E (expectation) step, the camera pose is fixed and the 3-D–2-D correspondences are established by Step 1 and Step 2 above. In

the M (maximization) step, the 3-D–2-D correspondences are fixed and the camera pose is updated by Step 3 above; (2) the novel approach has an advantage of easy implementation, since it does not need to extract any structural features from images [14], [21]–[23], [40], [42] or estimate the motion parameters of interest [30], [34] from the possible point matches before they are evaluated; and finally (3) the novel approach can produce encouraging projective registration results, as demonstrated in this paper.

While the existing registration algorithms [14], [40], [42] mainly use feature matching to establish and/or evaluate the possible point correspondences, the extraction of structural features is often sensitive to noise, occlusion, and appearance and disappearance of points. The rigid motion constraints based registration algorithm, however, takes a different strategy for registration. Instead of matching features, this strategy evolves rigid motion constraints, involving both the possible point matches and the camera pose parameters. While the existing registration methods often take for granted that the finally established correspondences are real, the novel approach can guarantee that in the end, the 3-D–2-D correspondences established by the traditional ICP criterion and evaluated by the rigid motion constraints represent *somewhat* real correspondences.

The goal of this paper is twofold: one is to study the behavior of point match quality measures [29] for the evaluation of 3-D–2-D point correspondences established by the traditional ICP criterion; the other is to justify both theoretically and experimentally the refined point matches. Theoretically, for the 3-D–2-D point matches subject to a single rigid translational motion and established by the traditional ICP criterion, once more than two of them satisfy the rigid motion constraints *exactly*, then all those of them that satisfy the rigid motion constraints *exactly* satisfy the motion equation and thus, must represent real correspondences without any ambiguity and the rigid translational motion that brings the transformed 3-D model into alignment with its projective image is then uniquely determined. A large number of experiments based on both synthetic data and real images reveal that in practice, the rigid motion constraints often render the false matches to be successfully eliminated, thus producing encouraging projective registration results.

The rest of this paper is structured as follows. The application of rigid motion constraints for the elimination of false matches is explained in Section II, a justification of the refined correspondences is made in Section III and the experimental results based on both synthetic data and real images are presented in Section IV. Finally, we discuss a number of issues relative to the proposed registration algorithm and draw some conclusions in Section V.

II. SMALL PURE TRANSLATIONAL MOTION–PROJECTIVE REGISTRATION ALGORITHM

Assuming that the camera undergoes a pure translational motion [17], [28], [29], then the relationship between a 3-D model point \mathbf{p} and its corresponding 2-D projective image point \mathbf{P}' can be represented as

$$z' \begin{pmatrix} \mathbf{P}' \\ 1 \end{pmatrix} = \mathbf{p} + \mathbf{t} \quad (1)$$

where the classical pinhole camera model is adopted and z' represents the depth of point \mathbf{p} in the camera centered coordinate frame and \mathbf{t} represents the position of the camera in the scene centered coordinate frame. For convenience of computation without loss of generality, we assume in this paper that the focal length of the camera is equal to 1. Assuming that the camera undergoes a small motion, the problem is thus, defined as the projective registration of free form surfaces subject to that small motion. The free form surfaces are represented as two sets of sparse points where one set contains the 3-D model data points $\mathbf{p}_i = (x_i, y_i, z_i)^T$ ($i = 1, 2, \dots, n_1 | n_1 \geq 2$) and another set contains the 2-D projective image data points $\mathbf{P}'_j = (X'_j, Y'_j)^T$ ($j = 1, 2, \dots, n_2 | n_2 \geq 2$). Assuming that two sets of points are overlapping in 3-D space, n_1 is not necessarily equal to n_2 due to occlusion and appearance and disappearance of points. The 3-D–2-D point pair with the same subscript does not mean that they represent a 3-D–2-D point correspondence.

In the rest of this paper, the following notations are used: capital letters represent points, vectors or matrices, lower case letters represent scalars, the capital letter \mathbf{I} denotes the identity matrix, superscript T denotes the transpose of a vector or a matrix, $|\cdot|$ denotes the absolute value of a scalar, and $\|\cdot\|$ denotes the Euclidean norm of a vector.

A. Quality Measurement of Possible 3-D–2-D Correspondences

Once the camera position $\mathbf{t} = (\Delta_x, \Delta_y, \Delta_z)^T$ has been initialized [29], the resulting set of 2-D projected image points $\bar{\mathbf{P}}' = \{\bar{\mathbf{P}}'_1, \bar{\mathbf{P}}'_2, \dots, \bar{\mathbf{P}}'_{n_1}\}$ of 3-D model data points $\{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_{n_1}\}$ on the image plane can be calculated by (1). Using the traditional ICP criterion [1], [3], [44], for each point $\bar{\mathbf{P}}'_i$ in set $\bar{\mathbf{P}}'$ its closest point $\tilde{\mathbf{P}}'_i = (\tilde{X}'_i, \tilde{Y}'_i)^T$ can be found in set $\mathbf{IP}' = \{\mathbf{P}'_1, \mathbf{P}'_2, \dots, \mathbf{P}'_{n_2}\}$

$$\tilde{\mathbf{P}}'_i = \operatorname{argmin}_{\mathbf{P}' \in \mathbf{IP}'} \|\mathbf{P}' - \bar{\mathbf{P}}'_i\| \quad (2)$$

which minimizes the Euclidean distance between the transformed point $\bar{\mathbf{P}}'_i$ and \mathbf{P}' in \mathbf{IP}' . The search space is determined by the size of the projective image \mathbf{IP}' . As a consequence of this operation, a set of possible 3-D–2-D point correspondences $(\mathbf{p}_i, \tilde{\mathbf{P}}'_i)$ ($i = 1, 2, \dots, n_1$) is obtained.

Unfortunately, the criterion used by the ICP algorithm essentially implies some ambiguities in determining 3-D–2-D point correspondences $(\mathbf{p}_i, \tilde{\mathbf{P}}'_i)$, since the closest point $\tilde{\mathbf{P}}'_i$ in the second frame to a transformed point $\bar{\mathbf{P}}'_i$ in the first frame is a necessary candidate of a correspondent, but not a sufficient due to the following facts: on the one hand, a single distance constraint cannot generally determine the exact position of a 2-D point on the image plane; on the other hand, some point matches established by the traditional ICP criterion are not feasible at all due to inaccurate initialization of camera pose, occlusion, appearance and disappearance of points, and noise presence in image data. As a result, additional techniques must be developed to eliminate false matches so that correct 3-D–2-D point correspondences can be established, leading the registration algorithm to converge more quickly and estimate more accurately the camera pose that aligns the two sets of data as well as possible. For this purpose, the quality of each

possible 3-D–2-D point correspondence $(\mathbf{p}_i, \tilde{\mathbf{P}}'_i)$ is defined as follows based on two rigid motion constraints (Fig. 1): one is about colinearity, the other is about equidistance.

From the focus of expansion (FOE) theory [18], it is known that if a point pair $(\mathbf{p}_i, \tilde{\mathbf{P}}'_i)$ represents a real 3-D–2-D correspondence and the camera position \mathbf{t} has been accurately initialized or estimated, then the image point $\tilde{\mathbf{P}}'_i = (\tilde{X}'_i, \tilde{Y}'_i)^T$ and the projections $\tilde{\mathbf{t}} = (\frac{\Delta_x}{\Delta_z}, \frac{\Delta_y}{\Delta_z})^T$ and $\tilde{\mathbf{p}}_i = (\frac{x_i}{z_i}, \frac{y_i}{z_i})^T$ of camera position \mathbf{t} and point \mathbf{p}_i on the image plane must be colinear (Fig. 1), and, thus, the following equality must hold:

$$\frac{\frac{\Delta_y}{\Delta_z} - \tilde{Y}'_i}{\frac{\Delta_x}{\Delta_z} - \tilde{X}'_i} = \frac{\frac{\Delta_y}{\Delta_z} - \frac{y_i}{z_i}}{\frac{\Delta_x}{\Delta_z} - \frac{x_i}{z_i}}. \quad (3)$$

However, in practice, due to inaccurate initial registration parameters, occlusion, appearance and disappearance of points, and noise distribution in image data, the point match $(\mathbf{p}_i, \tilde{\mathbf{P}}'_i)$ generally cannot satisfy this equation *exactly*. Thus, the relative difference between the two sides of this equation can then be used to measure the quality of the point match $(\mathbf{p}_i, \tilde{\mathbf{P}}'_i)$ based on which different 3-D–2-D point matches can be compared. The larger the difference between the slopes of line segments connecting $(\tilde{\mathbf{t}}, \tilde{\mathbf{P}}'_i)$ and $(\tilde{\mathbf{t}}, \tilde{\mathbf{p}}_i)$, the worse the point match $(\mathbf{p}_i, \tilde{\mathbf{P}}'_i)$. Thus based on the current registration parameter camera position \mathbf{t} , the quality of point match $(\mathbf{p}_i, \tilde{\mathbf{P}}'_i)$ can be defined as

$$g1_i = \left| \frac{\frac{\frac{\Delta_y}{\Delta_z} - \tilde{Y}'_i}{\frac{\Delta_x}{\Delta_z} - \tilde{X}'_i} - \frac{\frac{\Delta_y}{\Delta_z} - \frac{y_i}{z_i}}{\frac{\Delta_x}{\Delta_z} - \frac{x_i}{z_i}}}{\max \left(\left| \frac{\Delta_y}{\Delta_z} - \tilde{Y}'_i \right|, \left| \frac{\Delta_y}{\Delta_z} - \frac{y_i}{z_i} \right| \right)} \right|. \quad (4)$$

From (1), it is also known that

$$\|\mathbf{p}_i + 0.5\mathbf{t}\| = \|\tilde{\mathbf{p}}'_i - 0.5\mathbf{t}\| \quad (5)$$

where $\tilde{\mathbf{p}}'_i = \tilde{z}'_i(\tilde{X}'_i, \tilde{Y}'_i, 1)^T$ and the depth \tilde{z}'_i is estimated from the possible 3-D–2-D correspondence $(\mathbf{p}_i = (x_i, y_i, z_i)^T, \tilde{\mathbf{P}}'_i = (\tilde{X}'_i, \tilde{Y}'_i)^T)$ and the current camera position $\mathbf{t} = (\Delta_x, \Delta_y, \Delta_z)^T$ as: $\tilde{z}'_i = z_i + \Delta_z$. Equation (5) can be rewritten as: $\|-\mathbf{p}_i - 0.5\mathbf{t}\| = \|\tilde{\mathbf{p}}'_i - 0.5\mathbf{t}\|$ which shows that the reflection $-\mathbf{p}_i$ of point \mathbf{p}_i in 3-D space and the recovered 3-D point $\tilde{\mathbf{p}}'_i$ from the projective image point $\tilde{\mathbf{P}}'_i$ must be equidistant to the middle point of camera position \mathbf{t} (Fig. 1). Again, in practice, due to inaccurate initial registration parameters, occlusion, appearance and disappearance of points, and noise distribution in image data, the point match $(\mathbf{p}_i, \tilde{\mathbf{P}}'_i)$ generally cannot satisfy this equation *exactly*. Thus, based on the current registration parameter camera position \mathbf{t} , the quality of point match $(\mathbf{p}_i, \tilde{\mathbf{P}}'_i)$ based on this constraint can be similarly defined as

$$g2_i = \frac{\left| \|\mathbf{p}_i + 0.5\mathbf{t}\|^2 - \|\tilde{\mathbf{p}}'_i - 0.5\mathbf{t}\|^2 \right|}{\max \left(\|\mathbf{p}_i + 0.5\mathbf{t}\|^2, \|\tilde{\mathbf{p}}'_i - 0.5\mathbf{t}\|^2 \right)}. \quad (6)$$

B. False Match Elimination

Due to inaccurate initial registration parameters, occlusion, appearance and disappearance of points, and noise distribution in image data, the point matches $(\mathbf{p}_i, \tilde{\mathbf{P}}'_i)$ established by the

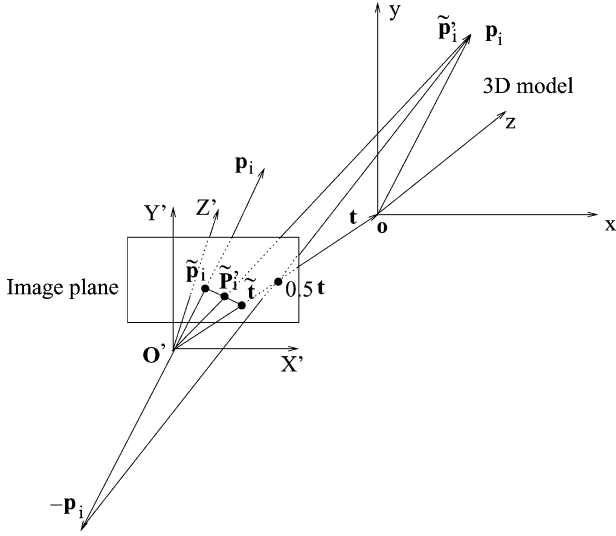


Fig. 1. Rigid motion constraints used. Image point $\bar{\mathbf{p}}'_i$ and the projections $\bar{\mathbf{p}}_i$ and $\bar{\mathbf{t}}$ of the 3-D model point \mathbf{p}_i and the camera position \mathbf{t} on the image plane are theoretically colinear. The recovered 3-D point $\tilde{\mathbf{p}}'_i$ from the image point $\bar{\mathbf{p}}'_i$ and the reflection $-\mathbf{p}_i$ of the 3-D model point \mathbf{p}_i are theoretically equidistant to the middle point $0.5\mathbf{t}$ of the camera position \mathbf{t} .

traditional ICP criterion represent just *pseudo* correspondences since, in the strict sense, at the beginning of registration, none of them can represent real correspondences. However, based on the point match quality measures defined above, relatively good point matches ($\mathbf{p}_i, \tilde{\mathbf{p}}'_i$) can be selected and used for camera position re-estimation. Eventually, accurate camera position \mathbf{t} that brings the transformed 3-D model points and its projective image points into alignment can be recovered. To this end, a statistical model to evaluate possible 3-D–2-D point correspondences ($\mathbf{p}_i, \tilde{\mathbf{p}}'_i$) is developed [29], [30], [34] as follows:

First the mean μ and standard deviation σ of the point match qualities $g1$ and $g2$ are computed as

$$\mu_{g1} = \frac{1}{n_1} \sum_{i=1}^{n_1} g1_i, \quad \sigma_{g1} = \sqrt{\frac{1}{n_1} \sum_{i=1}^{n_1} (g1_i - \mu_{g1})^2}$$

$$\mu_{g2} = \frac{1}{n_1} \sum_{i=1}^{n_1} g2_i, \quad \sigma_{g2} = \sqrt{\frac{1}{n_1} \sum_{i=1}^{n_1} (g2_i - \mu_{g2})^2}$$

Then the following rule is used to judge whether or not a specific 3-D–2-D correspondence ($\mathbf{p}_i, \tilde{\mathbf{p}}'_i$) is a false one: If either $|g1_i - \mu_{g1}| > \kappa \sigma_{g1}$ or $|g2_i - \mu_{g2}| > \kappa \sigma_{g2}$, then it is regarded as a false one. Otherwise, it is regarded as a feasible one. All feasible correspondences ($\mathbf{q}_i, \tilde{\mathbf{q}}'_i$) ($i = 1, 2, \dots, n | n \leq n_1$) among ($\mathbf{p}_i, \tilde{\mathbf{p}}'_i$) ($i = 1, 2, \dots, n_1$) are called the refined correspondences.

The parameter κ is user defined and controls how far from the mean a given relative difference must be for a point pair to be considered as a false match. It reflects not only the quantity of overlapping between a 3-D model and its projective image in 3-D space, but also the distribution of points. The rationale behind this rule is that the quality measures for different point matches can be considered as stochastic variables

$$G1_i = \frac{g1_i - \mu_{g1}}{\sigma_{g1}}, \quad G2_i = \frac{g2_i - \mu_{g2}}{\sigma_{g2}}.$$

The more distant either $g1_i$ or $g2_i$ is from the mean μ_{g1} or μ_{g2} , the more likely the corresponding point match ($\mathbf{p}_i, \tilde{\mathbf{p}}'_i$) represents a false match. As a result, as long as the parameter κ is properly defined, a set of refined correspondences ($\mathbf{q}_i, \tilde{\mathbf{q}}'_i$) ($i = 1, 2, \dots, n | n \leq n_1$) that are more likely to represent real correspondences is obtained based on which the camera position \mathbf{t} can be more accurately updated [29].

C. Pure Translational Motion- Projective Registration Algorithm

A novel algorithm for the projective registration of a free form surface subject to a pure translational motion is proposed as follows.

1. Initialize the pure translational motion [29], set iteration number $k = 0$.
2. Use the traditional ICP criterion (2) to establish a set of possible 3-D–2-D correspondences ($\mathbf{p}_i, \tilde{\mathbf{p}}'_i$) ($i = 1, 2, \dots, n_1$) between a 3-D model and its projective image.
3. Use (4) and (6) to compute the quality of possible 3-D–2-D correspondences ($\mathbf{p}_i, \tilde{\mathbf{p}}'_i$).
4. Use the statistical model described in Section II-B to eliminate false matches from the possible 3-D–2-D point correspondences ($\mathbf{p}_i, \tilde{\mathbf{p}}'_i$), leading to a set of refined correspondences ($\mathbf{q}_i, \tilde{\mathbf{q}}'_i$) ($i = 1, 2, \dots, n | n \leq n_1$).
5. Use the procedure described in [29] to update the camera position \mathbf{t} based on the refined correspondences ($\mathbf{q}_i, \tilde{\mathbf{q}}'_i$).
6. Apply the newly estimated camera position \mathbf{t} to each 3-D model point \mathbf{p}_i , yielding its projected image point $\bar{\mathbf{p}}'_i$ on the image plane.
7. When the variation of the camera position at two successive iterations is larger than the desired registration error ρ or when the iteration number k is larger than the maximum M , the algorithm terminates. Otherwise, $k \leftarrow k + 1$, go to step 2.

Parameters κ , ρ and M play a crucial role in characterizing algorithm performance. If κ is set too large, then a number of false matches will be introduced. If κ is set too small, then a number of good matches will be rejected. All these cases will bias the algorithm for the estimation of camera pose [9]. If ρ is set too large or M is set too small, then the algorithm can easily terminate pre-maturely. If ρ is set too small or M is set too large, then the algorithm requires intensive computation without a significant improvement to the existing projective registration results. Our experience has shown that when $\kappa \in [1.0, 2.0]$, $\rho \in [0.0001, 0.001]$, and $M \in [100, 400]$, good registration results can be obtained. In the experiments described below, we let $\kappa = 1.75$, $\rho = 0.0001$, and $M = 300$. The proposed algorithm is called the iterative closest point with the focus of expansion (ICPF) algorithm.

While different κ s do not necessarily yield more accurate projective registration results, the normalized relative quality measures $G1$ and $G2$ have nothing to do with the specific metric used to represent the quality measures $g1$ and $g2$. Since we have no reason to prefer one quality measure over the other, we use

the same κ for these independent quality measures as defined in this paper. As discussed and demonstrated in Sections III and IV, these two independent constraints used here are capable of accurately evaluating the possible 3-D–2-D correspondences established by the traditional ICP criterion.

III. THEORETICAL JUSTIFICATION OF THE REFINED CORRESPONDENCES

Since the refined correspondences $(\mathbf{q}_i, \tilde{\mathbf{Q}}'_i)$ are a subset of the ICP criterion established possible 3-D–2-D point matches $(\mathbf{p}_i, \tilde{\mathbf{P}}'_i)$, for the sake of brevity without introducing additional notations, we will use $(\mathbf{q}_i, \tilde{\mathbf{Q}}'_i)$ and $(\mathbf{p}_i, \tilde{\mathbf{P}}'_i)$ interchangeably in the rest of this section without causing any confusion in understanding the resulting contents.

A. Correspondence Definition

Before we justify the refined correspondences $(\mathbf{q}_i, \tilde{\mathbf{Q}}'_i)$, it is necessary to clarify what a 3-D–2-D correspondence here means. A real 3-D–2-D correspondence has two definitions: one is that given the camera pose parameters rigid rotation matrix \mathbf{R} and translation vector \mathbf{t} , a pair of a 3-D point \mathbf{p} and a 2-D point \mathbf{P}' satisfy the motion equation: $z'(\mathbf{P}') = \mathbf{R}\mathbf{p} + \mathbf{t}$, the other is that a pair of a 3-D and a 2-D point subject to a rigid motion (\mathbf{R}, \mathbf{t}) represent the same physical point in 3-D space. While feature matching based registration algorithms [14], [16], [32], [33], [40] often follow the second definition of a correspondence, applying invariant geometric or optical features from different images to represent the same physical points, our proposed method follows the first, evolving the rigid motion constraints and in the end, leading to that given the camera position \mathbf{t} , the refined correspondences $(\mathbf{q}_i, \tilde{\mathbf{Q}}'_i)$ satisfy the motion (1) with $\mathbf{R} = \mathbf{I}$.

B. Derivation

From the description of the algorithm in the last section, it is known that at the beginning of registration, since the motion parameters are not accurate, the relative differences between the two sides of the rigid motion constraints are large. As the registration progresses, even though the relative differences do not necessarily evolve monotonically, the overall trend is that they become smaller and smaller (this may occur in the last few iterations). Eventually, the differences relative to all real correspondences will be close to zero due to noise. This phenomenon has been demonstrated in Figs. 2 and 4.

Assuming that there are more than two 3-D–2-D point pairs $(\mathbf{p}_i, \tilde{\mathbf{P}}'_i)$ (where one 3-D point cannot occlude the other) whose quality measures as defined by (4) and (6) are zero at the end of registration, then we have: $|\frac{(\Delta_y/\Delta_z - \tilde{Y}'_i)/(\Delta_x/\Delta_z - \tilde{X}'_i)}{(\Delta_y/\Delta_z - y_i/z_i)/(\Delta_x/\Delta_z - x_i/z_i)}| / (\max(\Delta_y/\Delta_z - \tilde{Y}'_i)/(\Delta_x/\Delta_z - \tilde{X}'_i), (\Delta_y/\Delta_z - y_i/z_i)/(\Delta_x/\Delta_z - x_i/z_i)) = 0$ and $(\|\mathbf{p}_i + 0.5\mathbf{t}\|^2 - \|\tilde{\mathbf{p}}'_i - 0.5\mathbf{t}\|^2) / \max(\|\mathbf{p}_i + 0.5\mathbf{t}\|^2, \|\tilde{\mathbf{p}}'_i - 0.5\mathbf{t}\|^2) = 0$. Since in general the denominators of the fractions in these two equations are not zero, the numerators of the

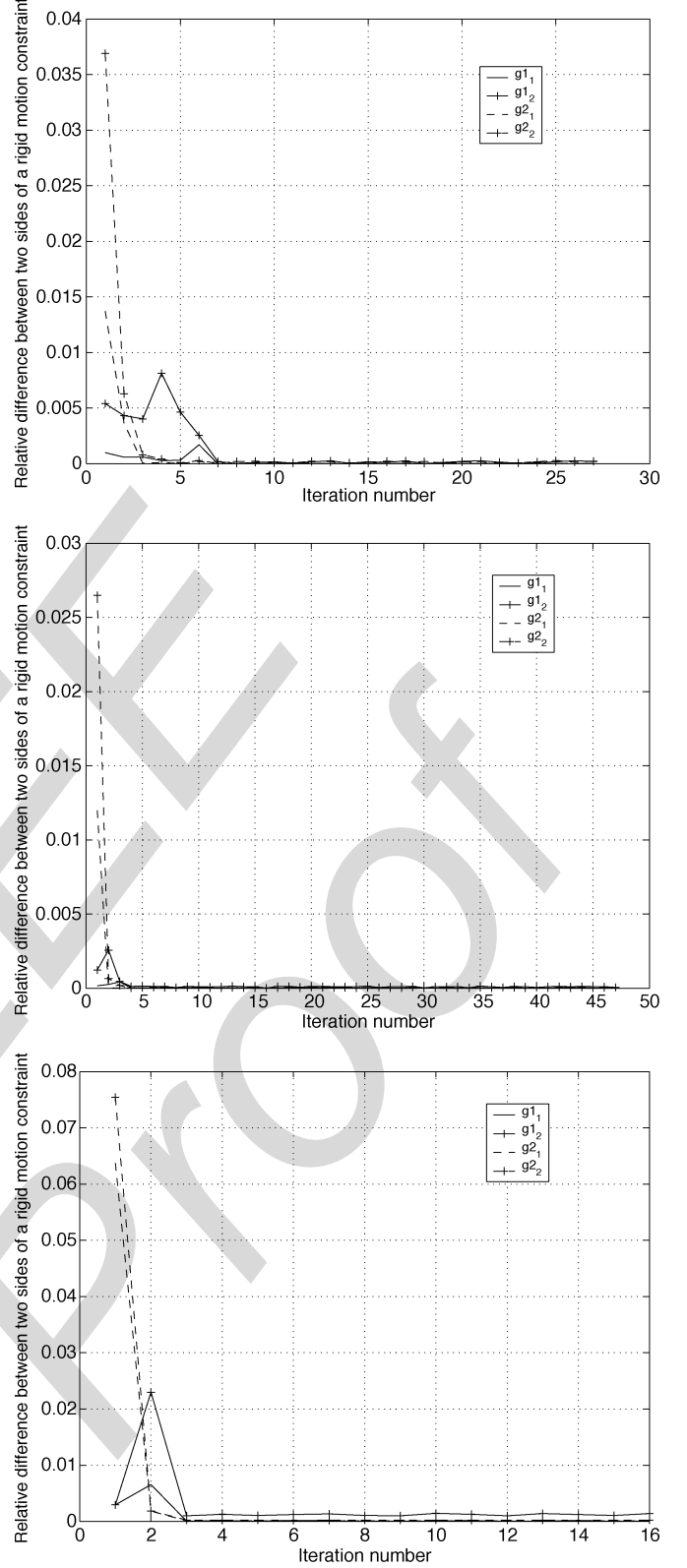


Fig. 2. Evolution of the two smallest relative differences $g1_1$, $g1_2$, $g2_1$, and $g2_2$ for point match quality measures $g1$ and $g2$, respectively, based on real images as described in Section IV. Top: Castle. Middle: Wall and tower. Bottom: Planar texture.

$$\frac{\frac{\Delta y}{\Delta z} - \tilde{Y}'_i}{\frac{\Delta x}{\Delta z} - \tilde{X}'_i} = \frac{\frac{\Delta y}{\Delta z} - \frac{y_i}{z_i}}{\frac{\Delta x}{\Delta z} - \frac{x_i}{z_i}} \quad (7)$$

$$\begin{pmatrix} \tilde{X}'_i \\ \tilde{Y}'_i \end{pmatrix} = a \begin{pmatrix} \frac{x_i}{z_i} \\ \frac{y_i}{z_i} \end{pmatrix} + (1-a) \begin{pmatrix} \frac{\Delta x}{\Delta z} \\ \frac{\Delta y}{\Delta z} \end{pmatrix}. \quad (9)$$

$$\begin{pmatrix} \tilde{X}'_i \\ \tilde{Y}'_i \\ 1 \end{pmatrix} = \frac{a}{z_i} \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} + \frac{1-a}{\Delta_z} \begin{pmatrix} \Delta_x \\ \Delta_y \\ \Delta_z \end{pmatrix} = \frac{a}{z_i} \mathbf{p}_i + \frac{1-a}{\Delta_z} \mathbf{t}. \quad (10)$$
$$\mathbf{p}_i^T \mathbf{p}_i + \mathbf{p}_i^T \mathbf{t} + 0.25 \mathbf{t}^T \mathbf{t} = \tilde{z}_i'^2 \begin{pmatrix} \tilde{X}_i' \\ \tilde{Y}_i' \\ 1 \end{pmatrix} - \tilde{z}_i' \mathbf{t}^T \begin{pmatrix} \tilde{X}_i' \\ \tilde{Y}_i' \\ 1 \end{pmatrix} + 0.25 \mathbf{t}^T \mathbf{t}. \quad (11)$$
$$\begin{aligned} & \frac{\tilde{z}_i'^2 a^2}{\tilde{z}_i'^2} \mathbf{p}_i^T \mathbf{p}_i + \left(\frac{2\tilde{z}_i'^2 a(1-a)}{z_i \Delta_z} - \frac{\tilde{z}_i' a}{z_i} \right) \mathbf{p}_i^T \mathbf{t} + \\ & \left(\frac{\tilde{z}_i'^2 (1-a)^2}{\Delta_z^2} - \frac{\tilde{z}_i' (1-a)}{\Delta_z} \right) \mathbf{t}^T \mathbf{t} = \mathbf{p}_i^T \mathbf{p}_i + \mathbf{p}_i^T \mathbf{t}. \quad (12) \end{aligned}$$


$$\frac{2\tilde{z}_i'^2 a(1-a)}{z_i \Delta_z} - \frac{\tilde{z}_i' a}{z_i} = 1 \quad (14)$$

$$\frac{\tilde{z}_i'^2(1-a)^2}{\Delta_z^2} - \frac{\tilde{z}_i'(1-a)}{\Delta_z} = 0. \quad (15)$$

$$\tilde{z}'_i = z_i + \Delta_z, \quad a = \frac{z_i}{z_i + \Delta_z}, \quad (16)$$

$$\tilde{z}'_i = -z_i + \Delta_z, \quad a = \frac{-z_i}{-z_i + \Delta_z}. \quad (17)$$

Substituting (16) into (10) results in

$$\begin{pmatrix} \tilde{X}'_i \\ \tilde{Y}'_i \\ 1 \end{pmatrix} = \frac{1}{z_i + \Delta_z} \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} + \frac{1}{z_i + \Delta_z} \begin{pmatrix} \Delta_x \\ \Delta_y \\ \Delta_z \end{pmatrix} \\ = \frac{1}{z_i + \Delta_z} \mathbf{p}_i + \frac{1}{z_i + \Delta_z} \mathbf{t} \quad (18)$$

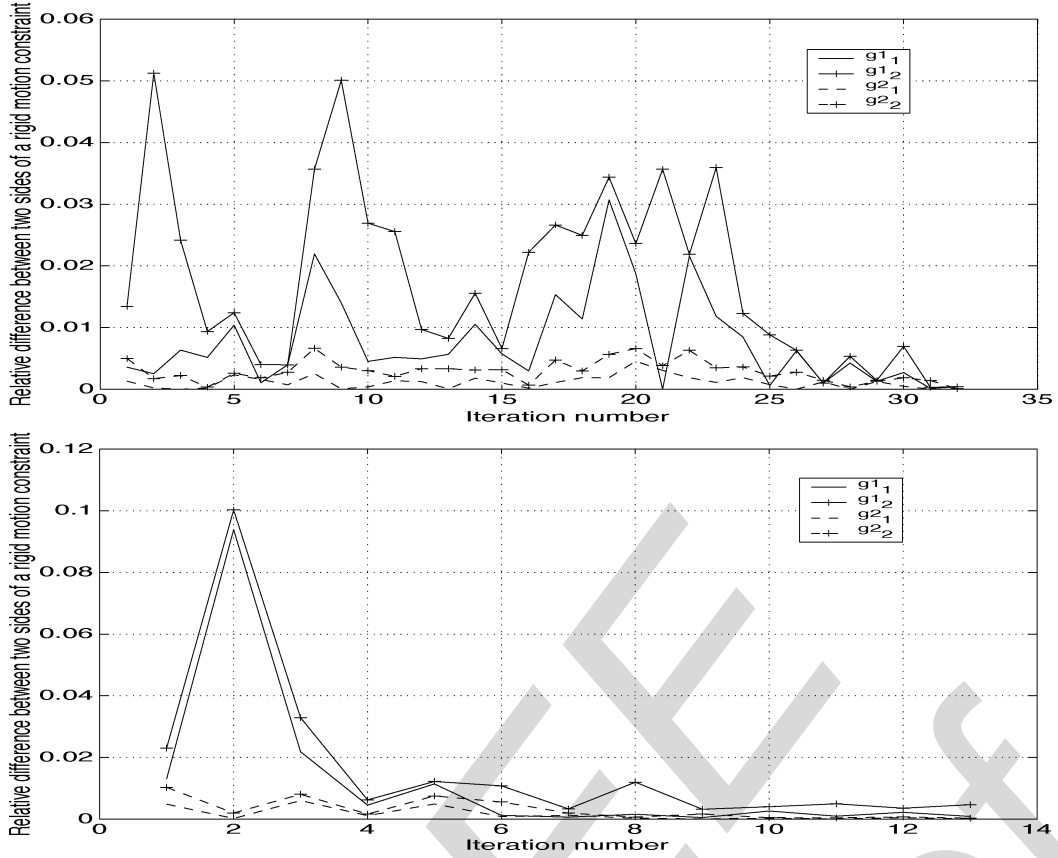


Fig. 4. Evolution of the two smallest relative differences g_{1_1} , g_{1_2} , g_{2_1} , and g_{2_2} for point match quality measures g_1 and g_2 respectively based on synthetic data as described in the next section. Top: $n = 100$. Bfottom: $n = 50$.

which is equivalent to

$$(z_i + \Delta_z) \begin{pmatrix} \tilde{X}'_i \\ \tilde{Y}'_i \\ 1 \end{pmatrix} = \mathbf{p}_i + \mathbf{t}. \quad (19)$$

This shows that given camera position \mathbf{t} , for the point pairs $(\mathbf{p}_i, \tilde{\mathbf{P}}'_i)$ established by the ICP criterion, as long as at least two of them satisfy the rigid motion constraints *exactly*, then all those of them that satisfy the rigid motion constraints *exactly* satisfy the motion (1) and thus from the first definition of a correspondence above, it is known that they must represent real correspondences without any ambiguity. In this case, for any 3-D model point \mathbf{p}_i , once the camera position \mathbf{t} is given, then the image point $\tilde{\mathbf{P}}'_i$ selected by the ICP criterion that satisfies the rigid motion constraints exactly is uniquely determined and must be the real correspondent of \mathbf{p}_i .

Note that we made a strong assumption here that more than two 3-D–2-D point pairs satisfy the two rigid motion constraints *exactly*. As demonstrated in Fig. 4, in practice, due to noise presence in real data, these two rigid motion constraints are often violated. The extent of violation depends on the distribution of points and noise level corrupting data points. However, this does not affect deepening our understanding of the conditions under which accurate 3-D–2-D point correspondences can be established.

The above analysis shows that evolving the rigid motion constraints is very attractive for the projective registration of a 3-D model and its projective image. The possible 3-D–2-D

point matches established by the ICP criterion and evaluated by the rigid motion constraints will all *approximately* satisfy the motion (1), due to noise and some poor point matches and thus, it is guaranteed that from the first definition of a correspondence above, they represent *somewhat* real correspondences, achieving encouraging projective registration results. The experiments based on both synthetic data and real images described in the next section verify our claim here.

IV. EXPERIMENTAL RESULTS

After we theoretically justified the refined 3-D–2-D point correspondences in the last section, in this section, we validate them experimentally based on both synthetic data and real images, simulating or representing various imaging scenarios in the real world. The ICPF algorithm was designed to be applicable to sparse points. In practice, sparse points can be selected either manually or automatically [2], [11], [45]. For a feeling of noise level corrupting the image data and for a comparative study of performance, we also implemented the focus of expansion and least squares (FLS) algorithm [29] based on all the given correspondences.

In contrast with the assumption that all image points can find correspondents on the 3-D model [28], [43], we considered in this paper a general scenario that the 3-D models and their projective images include disappearing and appearing points respectively. This implies that in practice, no constraints are imposed on the 3-D model construction or on the projective image

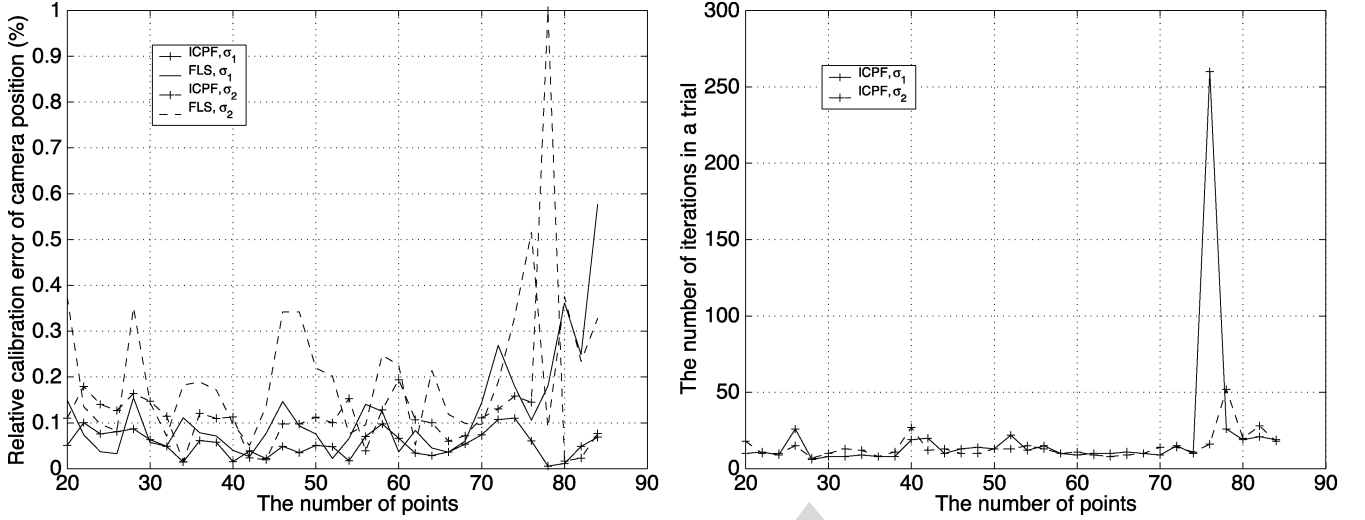


Fig. 5. Relationship between (left) the relative calibration error of estimated camera position and (right) the number of iterations in a trial and the number of points on a free form surface.

acquisition. Consequently, the ICPF algorithm facilitates further development and can easily be modified for real applications.

For the simulation of appearance and disappearance of points in the real world data, we varied the percentages from 0% to 20%, corresponding to small motions, of disappearing and appearing points in the 3-D models and their projective images respectively. A significant difference between the FLS and ICPF algorithms lies in that while the former requires the knowledge of 3-D–2-D point correspondences, the latter does not and it only requires that the two sets of points are overlapping in 3-D space and the overlapping is relatively large (70%, for example).

The reason why the FLS algorithm is chosen is that most of the existing registration algorithms [9], [24] are iterative, involving a number of parameters to be set by the user or require human intervention [27] or feature extraction [14], [21]–[23], [40], [42]. If the parameters are set properly or the structural features are accurately extracted, then the algorithms are accurate. Otherwise, they can degenerate catastrophically. Thus, a comparative study between these algorithms and the proposed one is more or less difficult. The FLS algorithm, however, has a closed form solution without requiring any human intervention and, thus, can provide an objective reference for the performance assessment of the ICPF algorithm. The parameters of interest in this paper are: the two smallest relative differences g_{11} , g_{12} , g_{21} , and g_{22} for each point match quality measure at different iterations and the relative calibration error e_t of estimated camera position $\hat{\mathbf{t}}$. In Figs. 2 and 4, solid lines correspond to the point match quality measure g_1 , dash lines correspond to the point match quality measure g_2 , lines without any signs correspond to the smallest relative difference, lines with plus signs correspond to the second smallest relative difference. For a better visualization of projective registration results, the two sets of transformed 3-D model and 2-D projective image points at iteration 1, in the middle of registration process and after registration are also presented in Figs. 6, 9, 10, and 12. In these figures, circles represent projective image points, plus signs represent the transformed 3-D model points.

TABLE I
AVERAGE μ AND STANDARD DEVIATION σ OF THE RELATIVE CALIBRATION ERROR e_t IN PERCENTAGE OF THE ESTIMATED CAMERA POSITION $\hat{\mathbf{t}}$ FOR THE FLS AND ICPF ALGORITHMS AND THE ITERATION NUMBERS k IN DIFFERENT TRIALS FOR THE ICPF ALGORITHM USING SYNTHETIC DATA CORRUPTED BY DIFFERENT LEVELS OF NOISE

Noise	Measure	Method	e_t (%)	k
σ_1	μ	ICPF	0.05	20.61
		FLS	0.12	
	σ	ICPF	0.03	42.65
		FLS	0.11	
σ_2	μ	ICPF	0.13	14.42
		FLS	0.21	
	σ	ICPF	0.13	8.14
		FLS	0.16	

A. Synthetic Points Data

In this section, we present the experimental results based on synthetic points data which can provide an objective evaluation of the proposed algorithm. A set of n points \mathbf{p}_i were first randomly generated with uniform distribution within $[-100\ 100] \times [-100\ 100] \times [-100\ 100]$ which were then subject to a translational motion \mathbf{t} randomly generated with uniform distribution within $[100\ 200] \times [100\ 200] \times [100\ 200]$. Finally we projected these transformed points onto the image plane yielding their corresponding image points \mathbf{P}'_i (care has to be taken to ensure that all transformed points are in front of the camera). Thus, we have precise knowledge about the 3-D–2-D correspondences $(\mathbf{p}_i, \mathbf{P}'_i)$ and the camera position \mathbf{t} which serve as reference for the performance assessment of different algorithms. In order to simulate the noise corrupted real world data, zero-mean Gaussian random noise with standard deviation $\sigma = 0.005$ was added to the coordinates of each 3-D model and 2-D projective image point respectively in one series of experiments, $\sigma = 0.01$ in another. While the FLS algorithm was applied to all correspondence data, the ICPF algorithm was applied to the data, unless otherwise stated, with the last 10% 3-D model points ($n_1 = 0.9n$) and the beginning 20% projective image points

($n_2 = 0.8n$) removed respectively for the simulation of appearance and disappearance of points.

1) *Different Point Sets With the Same Percentage of Appearing and Disappearing Points:* In this section, we test our algorithm based on different synthetic points data with fixed percentages of appearing and disappearing points. This in practice simulates such a scenario that the camera has a fixed field of view at the same position. The experimental results are presented in Figs. 4, 5, and 6 and Table I. In Fig. 5, the calibration error e_t was capped at 1% for better visualization.

From Fig. 5 and Table I, it can be seen that both the ICPF and FLS algorithms produce fluctuating results. This shows that the projective registration and motion estimation is sensitive to the configuration of points and the noise contaminating the data points. The ICPF algorithm is more accurate than the FLS algorithm. For example, while the FLS algorithm calibrated the average of the relative errors of camera position to be 0.21% with the synthetic data corrupted by heavier noise σ_2 , the ICPF algorithm calibrated the same parameter to be 0.13% with the synthetic data including 30% appearing and disappearing points. This is because while the FLS algorithm equally treated different 3-D–2-D correspondences despite the fact that they were corrupted by different levels of noise, the rigid motion constraints rejected not only false matches due to appearance and disappearance of points, but the noise highly corrupted 3-D–2-D correspondences as well. As a result, the ICPF algorithm estimated the camera pose using high quality 3-D–2-D correspondences, thus, achieving more accurate projective registration results. However, from Fig. 4, it can be seen that the rigid motion constraints were often violated since even the smallest two relative differences are not equal to zero. This shows that the rigid motion constraints provide sufficient, but not necessary, conditions in practice for a pair of a 3-D and a 2-D point to represent a real correspondence. The extent of violation depends on the distribution of points and noise level corrupting data points.

From Fig. 5 and Table I, it can also be seen that when the data are corrupted by higher level of noise, the ICPF algorithm produces worse results. This is expected since when the data points are corrupted by higher level of noise, the points are more unreliable. Even so, the ICPF algorithm still successfully established correct 3-D–2-D point matches, leading to accurate projective registration results.

From Fig. 6, it can be seen that at the beginning of projective registration, the two sets of projective image and transformed 3-D model points are significantly different on the image plane. With the registration progressing, the camera position becomes more and more accurate, leading the transformed 3-D model points to approach the projective image points as illustrated in the middle row of Fig. 6. Eventually, the two sets of transformed 3-D model and projective image points are perfectly superimposed in the image plane and appearing and disappearing points have been correctly identified. In this case, the ICPF algorithm successfully eliminated 30% false matches due to appearance and disappearance of points. This visualization clearly delivers the quality of the proposed ICPF algorithm.

2) *Same Point Sets With Different Percentages of Appearing and Disappearing Points:* In this section, we test our algorithm in dealing with different percentages of appearing and disap-

pearing points. This in practice simulates such a scenario that the camera has various fields of view at the same position. The 3-D model and its projective image initially included $n = 50$ points respectively. All the points were corrupted by lower level σ_1 of Gaussian noise described above. The percentages of appearing and disappearing points were then varied respectively from 0% to 20%, corresponding to small motions. The experimental results are presented in Table II.

From Table II, it can be seen that the ICPF algorithm produced stable results despite the fact that the data points include different percentages of appearing and disappearing points. In this case, the ICPF algorithm successfully eliminated up to 40% false matches. The larger the percentages of appearing and disappearing points, the worse the projective registration results. This is expected since more appearing and disappearing points present in the 3-D model and projective image render it difficult for the ICPF algorithm to distinguish feasible 3-D–2-D point matches from false ones.

From Table II, it is interesting to note that a smaller percentage of appearance and disappearance of points does not always imply that the ICPF algorithm will produce better results. For example, while the ICPF algorithm calibrated the relative error of camera position to be 6.25% with the synthetic data including 5% disappearing 3-D model points and 15% appearing projective image points, it calibrated the same parameter to be 0.03% with the synthetic data including 20% disappearing 3-D model and appearing projective image points respectively. This is because the final motion estimation is determined by the refined 3-D–2-D point matches. The higher the quality of the refined 3-D–2-D point matches, the better the final motion estimation results. The images with a smaller percentage of appearance and disappearance of points do not always imply that the finally established 3-D–2-D correspondences are of higher quality, leading to better projective registration results. The quality of the refined 3-D–2-D correspondences is mainly determined by κ , the distribution of points and the initialization of camera position and also effected by the Monte Carlo resampling technique, employed in the process of camera position update.

3) *Different Initializations:* In the above sections we fixed the initialization of the z coordinate of the camera position. In this section, we test whether the proposed ICPF algorithm is sensitive to this initialization. For this purpose, we ran the program 100 times and each time, the z coordinate was set as: $\Delta_z^{(0)} = (1 + 0.4\alpha)\Delta_z$ where Δ_z is the true value of the camera position and α is randomly generated with uniform distribution in the interval $[-1, 1]$. This in practice simulates such a scenario that the initialization of the z coordinate of the camera position can have as large as an error of 40%. The final error of the camera position calibration is measured as the median of 100 relative calibration errors e_t of the estimated camera position \hat{t} obtained for each trial [39]. The experimental results are presented in Fig. 7 and Table III. In Fig. 7, the calibration error e_t was capped again at 1% for better visualization.

From Fig. 7 and Table III, it can be seen that the ICPF algorithm produces stable results for the projective registration of a 3-D model and its projective image. This shows that the ICPF algorithm has a large region for convergence. Comparing Fig. 7

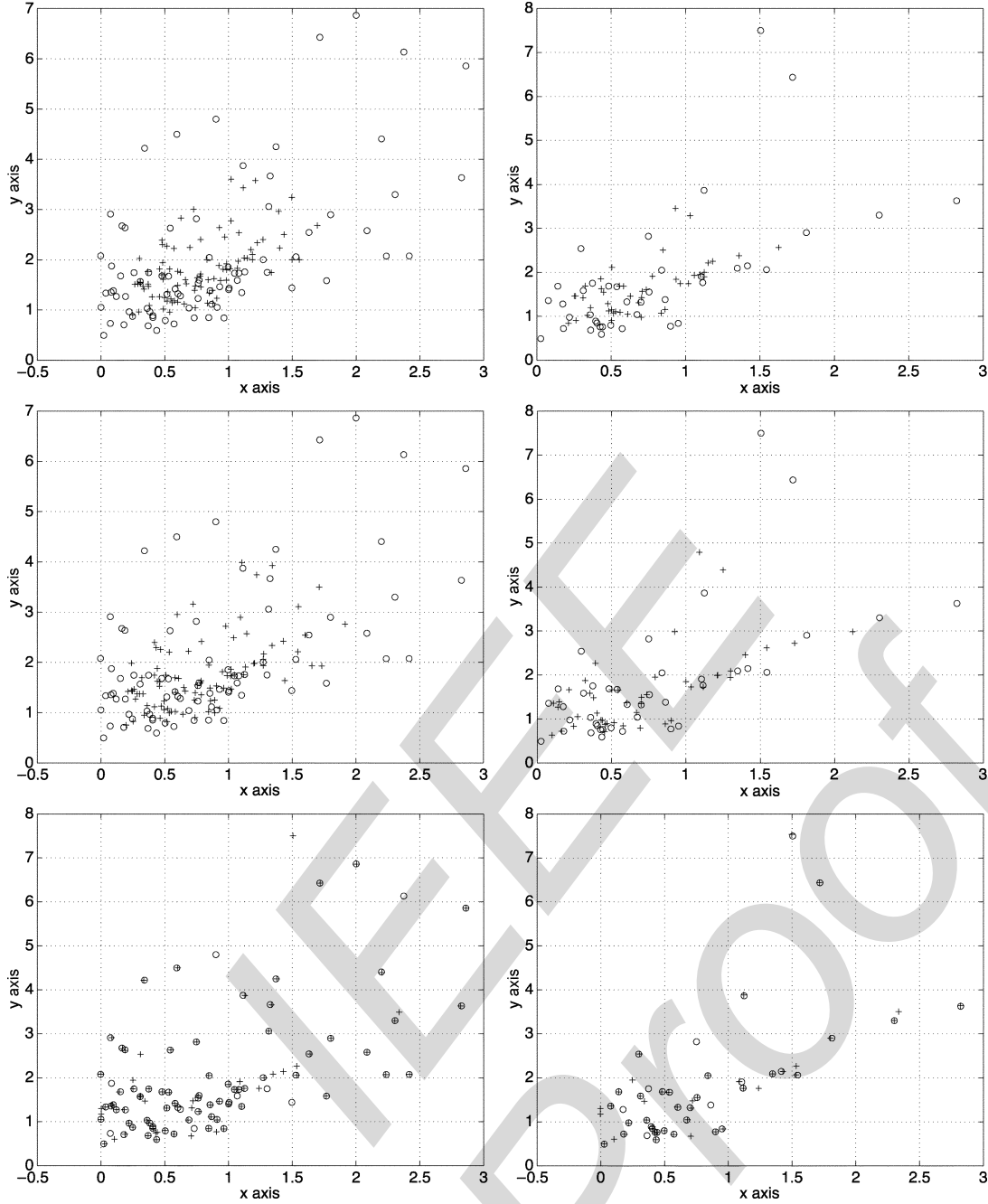


Fig. 6. Projective registration results based on synthetic points data. Top row: Result at iteration 1. Middle row: Result at the middle of registration. Bottom row: Result after registration. Left column: $n = 100$. Right column: $n = 50$.

TABLE II
RELATIVE CALIBRATION ERROR e_t IN PERCENTAGE OF THE CAMERA POSITION \hat{t} FOR THE ICPF ALGORITHM USING SYNTHETIC POINTS DATA WITH DIFFERENT PERCENTAGES OF APPEARING AND DISAPPEARING POINTS. COLUMN CORRESPONDS TO THE PERCENTAGE OF DISAPPEARING 3-D MODEL POINTS. ROW CORRESPONDS TO THE PERCENTAGE OF APPEARING PROJECTIVE IMAGE POINTS

$e_t(\%)$	0%	5%	10%	15%	20%
0%	0.04	0.05	0.07	0.09	0.03
5%	0.05	0.06	0.07	0.05	0.03
10%	0.06	0.06	0.06	0.10	0.05
15%	0.05	6.25	0.05	0.06	0.01
20%	0.05	6.56	0.10	0.09	0.03

with Fig. 5 and Table III with Table I, it can be seen that the algorithm performance improvement due to varied initializations is not significant. This is because the final registration results are a function of not just initialization, but a number of factors, including initialization, percentages of appearance and disappearance of points, κ , distribution of points, and noise level corrupting data points. In addition, while the camera position calibration error obtained by the ICPF algorithm with a fixed initialization $\Delta_z^{(0)}$ for the case $n = 78$ was reduced from 8.68% to 0.02% by the ICPF algorithm with varied initializations $\Delta_z^{(0)}$, the latter increased the camera position calibration error from 0.06% to 6.21% for the case $n = 76$. Thus, it can be observed that while the algorithm with various initializations solves one

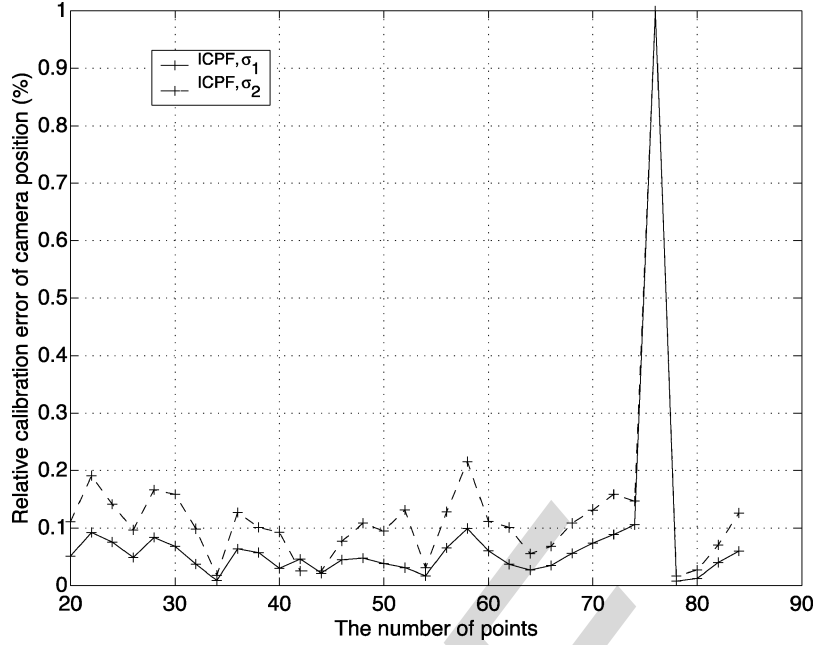


Fig. 7. Relationship between the relative calibration error of camera position and the number of points for the ICPF algorithm based on different initializations.

TABLE III
AVERAGE μ AND STANDARD DEVIATION σ OF THE RELATIVE CALIBRATION ERROR e_t IN PERCENTAGE OF THE ESTIMATED CAMERA POSITION \hat{t} FOR THE ICPF ALGORITHMS USING SYNTHETIC DATA CORRUPTED BY DIFFERENT LEVELS OF NOISES AND DIFFERENT INITIALIZATIONS

Noise	Measure	$e_t(\%)$
σ_1	μ	0.08
	σ	0.16
σ_2	μ	0.13
	σ	0.16

problem, it introduces another. This shows that varying initializations for the ICPF algorithm to improve performance is not always reliable despite the fact that it took 100 times the time the ICPF algorithm with a fixed initialization $\Delta_z^{(0)}$ used for projective registration. This phenomenon is usually also true for other registration algorithms [41]. Therefore, the algorithm itself should be accurate for good projective registration results.

Note that we use the same κ for all the experiments presented in this paper. We believe that if κ is set through considering the distribution of points, better results can be obtained. This is our research topic in the future.

B. Real Images

In this section, we test our algorithm based on real images. All the real images used in this paper were downloaded from the image database of Calibrated Imaging Laboratory at Carnegie Mellon University. Again, while the FLS algorithm was applied to all correspondence data, the ICPF algorithm was applied to the data, unless otherwise stated, with the last 10% 3-D model points ($n_1 = 0.9n$) and the beginning 20% projective image points ($n_2 = 0.8n$) removed respectively for both the castle and wall and tower images and with 10% of the last 3-D model points as well as the beginning projective image points ($n_1 = n_2 = 0.9n$) removed respectively for the planar texture image.

1) *Different Objects With Different Camera Positions:* In this section, we test our algorithm based on real images of different objects with different camera positions. This in practice simulates such a scenario that the camera has a fixed field of view at different positions. This series of experiments assesses the performance of the ICPF algorithm in dealing with various kinds of objects undergoing different small translational motions. The real images chosen are the castle, wall and tower, and planar texture images as depicted in Fig. 8. The experimental results are presented in Figs. 2, 9, and 10 and Table IV.

From Figs. 9 and 10, it can again be seen that at the beginning of registration, the two sets of transformed 3-D model and projective image points are significantly different on the image plane. In the middle of registration process, the two sets of points are already perfectly overlapping. Within the second half of registration time, the algorithm just refines the registration results, leading to a slight improvement to the registration results obtained in the first half of registration time. This visualization clearly shows that the proposed algorithm is accurate for the registration of a 3-D model and its projective image despite the fact that the actual z coordinate of the camera position is far away from its initialization of 120 [29]. In these cases, the algorithm successfully eliminated up to 30% false matches. From Fig. 2, however, it can be seen that there are no two pairs of 3-D–2-D point matches satisfying the rigid motion constraints exactly. But this does not materially affect the encouraging projective registration of real images used in this paper as confirmed by Table IV.

From Table IV, it can be seen that even though data used for experiments were optimized in the least squares sense [38], the FLS algorithm still calibrated the relative error of camera position to be 2.85% for the registration of the castle images and 1.08% for the registration of the planar texture images. This shows that the real image data are very noisy. Even so,

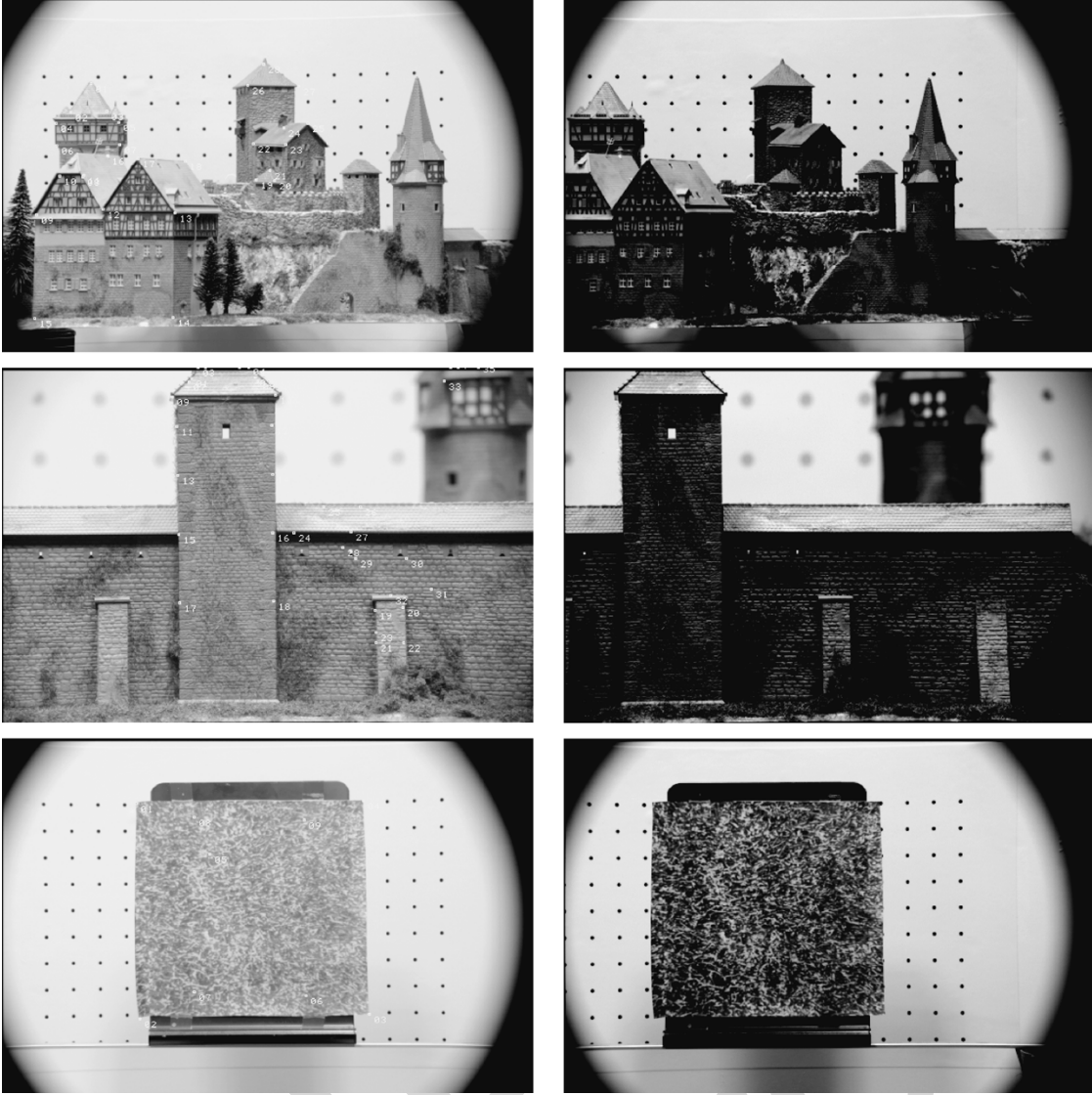


Fig. 8. Real images used. Left column: 3-D model. Right column: Projective image. Tow row: Castle1. Middle row: Wall and tower. Bottom row: Planar texture.

the ICPF algorithm still successfully registered the castle images with the relative calibration error of camera position being 3.05%, the wall and tower images with the relative calibration error of camera position being 0.24%, and the planar texture images with the relative calibration error of camera position being 1.28%. The reason why the ICPF algorithm yields a slightly worse registration results than the FLS algorithm is that the former had to deal with up to 30% false matches and applied only at most 70% feasible point correspondences in the overlapping area for camera position estimation. The feasible 3-D–2-D correspondences established by the ICPF algorithm were not necessarily optimal in the least squares sense, thus, producing slightly worse results. Recalling that the ICPF algorithm did not possess any knowledge about appearance and disappearance of points, the distribution of points, and noise corrupting data points and uniformly set the crucial parameter κ , it can be fairly concluded that the rigid motion constraints are powerful in evaluating the possible 3-D–2-D point correspondences established by the traditional ICP criterion.

2) *Same Object With Different Camera Positions:* In this section, we test our algorithm based on real images of the same object with different camera positions. This means that the 3-D model is the same, but the projective images were captured at different positions. This series of experiments provides a more objective evaluation of the robustness of the proposed algorithm as the experiments were based on the same object structure. The images chosen were the real castle images as depicted in Fig. 11. The experimental results are presented in Fig. 12 and Table V.

From Fig. 12 it can be seen that the ICPF algorithm exhibits a robust behavior in registering a 3-D model and its projective images captured at different positions. Even though the images were captured at different positions, the ICPF algorithm in this case always perfectly registered the two sets of 3-D model and projective image points and all appearing and disappearing points have been always correctly identified. Fig. 9 and Table V show that while the ICPF algorithm successfully eliminated 30% false matches, it produced accurate registration results with the relative calibration errors of camera positions around

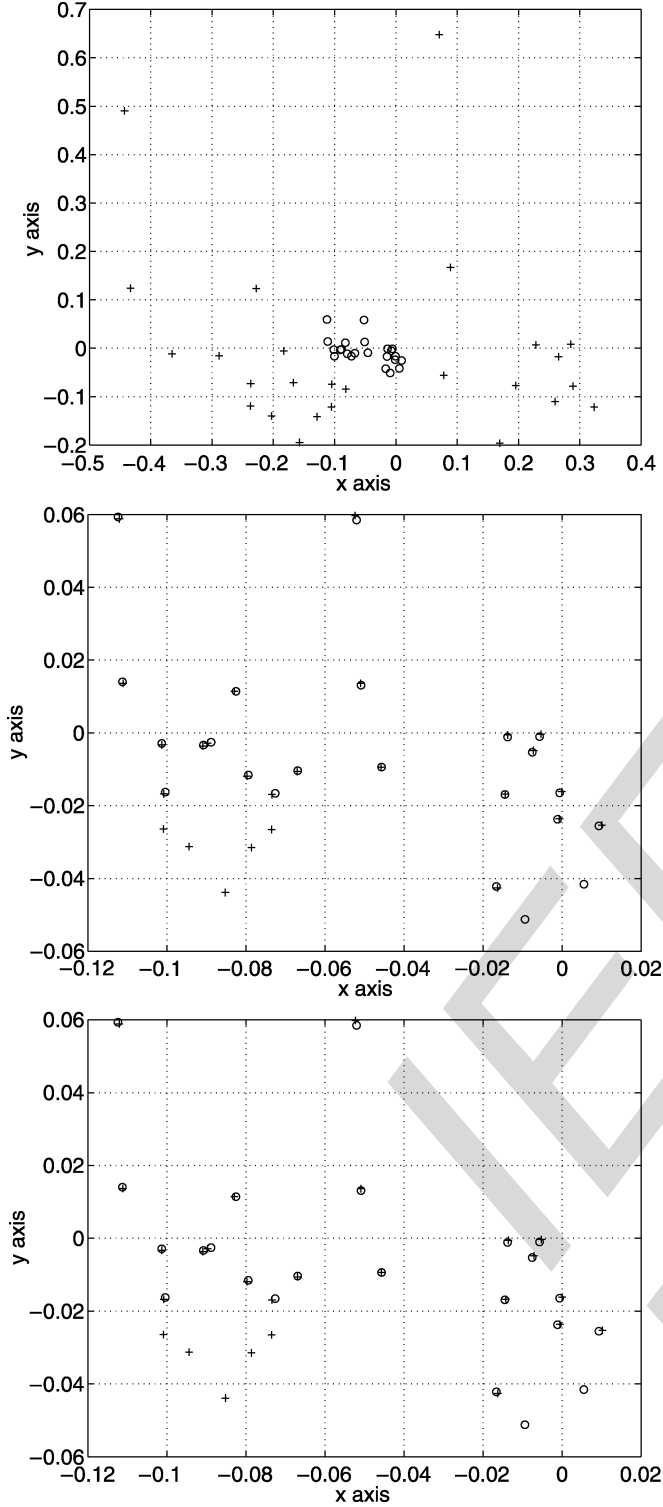


Fig. 9. Projective registration results based on real castle images. Top: Result at iteration 1. Middle: Result at the middle of registration. Bottom: Result after registration.

2.50%, which is comparable to those obtained by the FLS algorithm with given 3-D–2-D correspondences optimized in the least squares sense.

3) *Same Object With Different Percentages of Appearing and Disappearing Points:* In this section, we test our algorithm based on real images with different percentages from

0% to 20% of appearing and disappearing points. The images chosen are the castle and castle1 and wall and tower as depicted in Figs. 8 and 11. The experimental results are presented in Tables VI and VII.

From Tables VI and VII, it can again be seen that even though the ICPF algorithm did not possess any knowledge about appearance and disappearance of points, the distribution of points, and noise level corrupting image data and uniformly set the crucial parameter κ , it successfully handled up to 40% false matches established by the traditional ICP criterion due to appearance and disappearance of points, yielding very good registration results with relative calibration errors of camera positions around 3.00% for the registration of the castle images and 0.70% for the registration of the wall and tower images. This confirms that the ICPF algorithm is powerful for the projective registration of a 3-D model and its projective image with various percentages of appearance and disappearance of points.

4) *Different Initializations:* In this section, we test the proposed ICPF algorithm based on real images using various initializations. The initialization setup was the same as that described in Section IV-A-3. Then we ran the program 200 times and the final error of the camera position calibration is measured as the median of 200 relative calibration errors e_t of the estimated camera position $\hat{\mathbf{t}}$ obtained for each trial [39]. The experimental results are presented in Table VIII.

From Table VIII, it can be seen that the ICPF algorithm always correctly converged to the right solution, leading to very accurate projective registration results. This again shows that the ICPF algorithm has a large region for convergence. Comparing Table VIII with Tables V and IV, it can be seen that even though 200 experiments with different initializations were carried out respectively, the final results are just slightly better despite the fact that the ICPF algorithm with varied initializations $\Delta_z^{(0)}$ took 200 times the time the ICPF algorithm with a fixed initialization $\Delta_z^{(0)}$ used for projective registration. This phenomenon again stresses that the algorithm itself should be accurate and robust. The performance improvement due to repeating the experiments with various configurations of relative parameters is either small or time intensive.

All the experiments presented in this paper were implemented on a Pentium III, 256 MB RAM computer. The proposed algorithm converged within 3 seconds for any experiment described in this paper without optimizing the program code.

V. DISCUSSION AND CONCLUSIONS

A. Discussion

Through the experiments based on both synthetic data and real images, the following observations have been made.

1. Even though the assumption made in this paper that the camera underwent a pure translational motion is restrictive, the work reported is still significant in at least two aspects: firstly, we have gained a deep understanding of the conditions under which accurate 3-D–2-D correspondences can be established. These conditions can be applied to validate 3-D–2-D point correspondences established either by the traditional

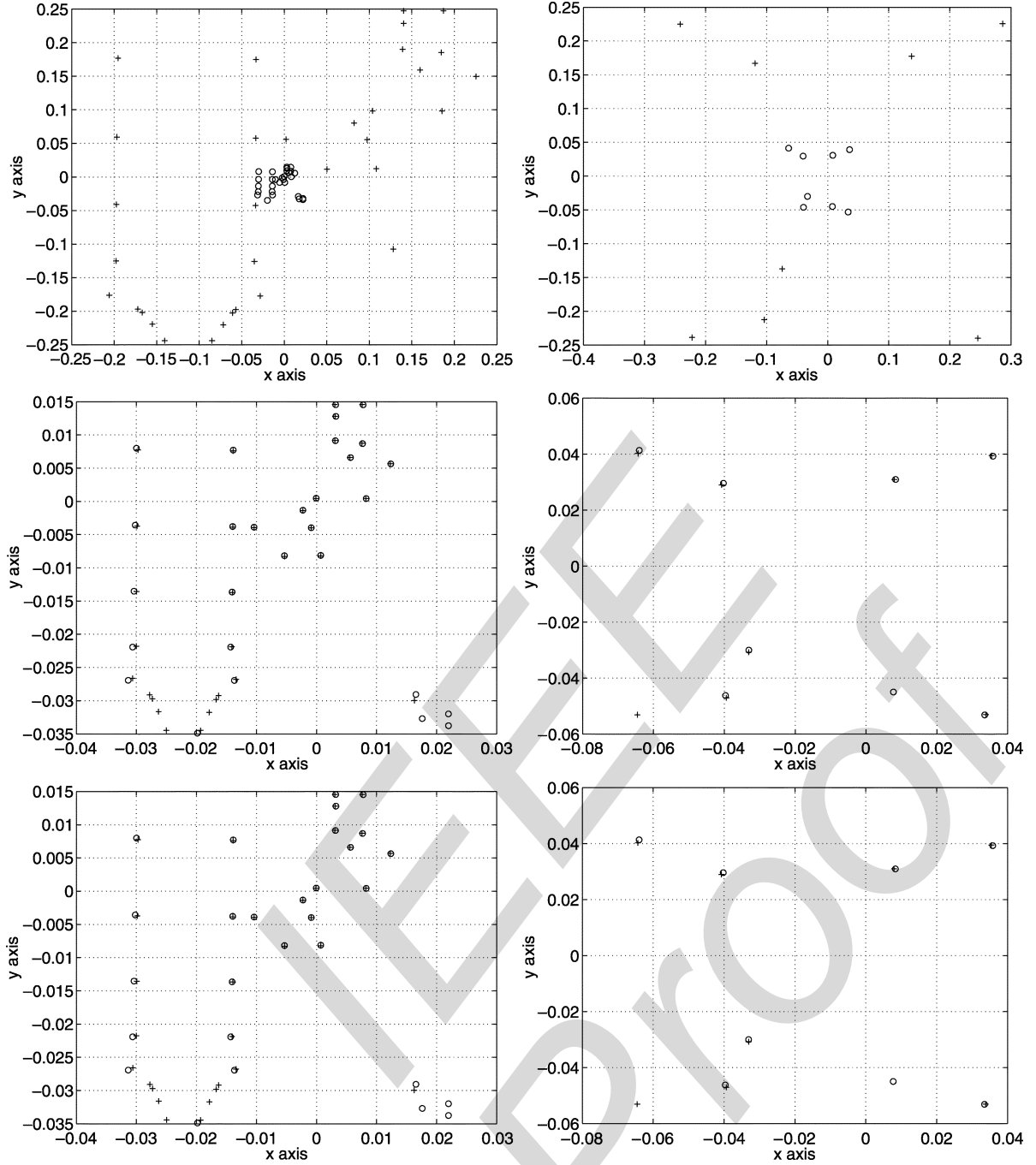


Fig. 10. Projective registration results based on real images. Top row: Result at iteration 1. Middle row: Result at the middle of registration. Bottom row: Result after registration. Left column: Wall and tower. Right column: Planar texture.

ICP criterion as demonstrated in this paper or by feature matching [31]; secondly, the theory and algorithm proposed provide a good starting point to develop algorithms for real applications with cameras subject to general motions, including both translational and rotational components. In this case, the camera motion is represented as: $z'(P'_1) = \mathbf{R}\mathbf{p} + \mathbf{t}$. If we denote $\mathbf{R}\mathbf{p} = \mathbf{q}$, then this equation can be rewritten as: $z'(P'_1) = \mathbf{q} + \mathbf{t}$. This shows that both the theory and algorithm proposed in this paper can then be used to register a rotated 3-D model and its projective image.

Thus, the extension of the proposed ICPF algorithm to accommodate the general motion is feasible.

2. Currently, we uniformly set κ without taking into account the distribution of points or concrete applications. We believe that better projective registration results can be achieved through dynamically setting κ , considering the distribution of points and concrete applications. Our experience has shown that under typical imaging conditions with overlapping relatively large (e.g., 70%) without any knowledge about occlusion, or appearance and disappearance of points, when κ

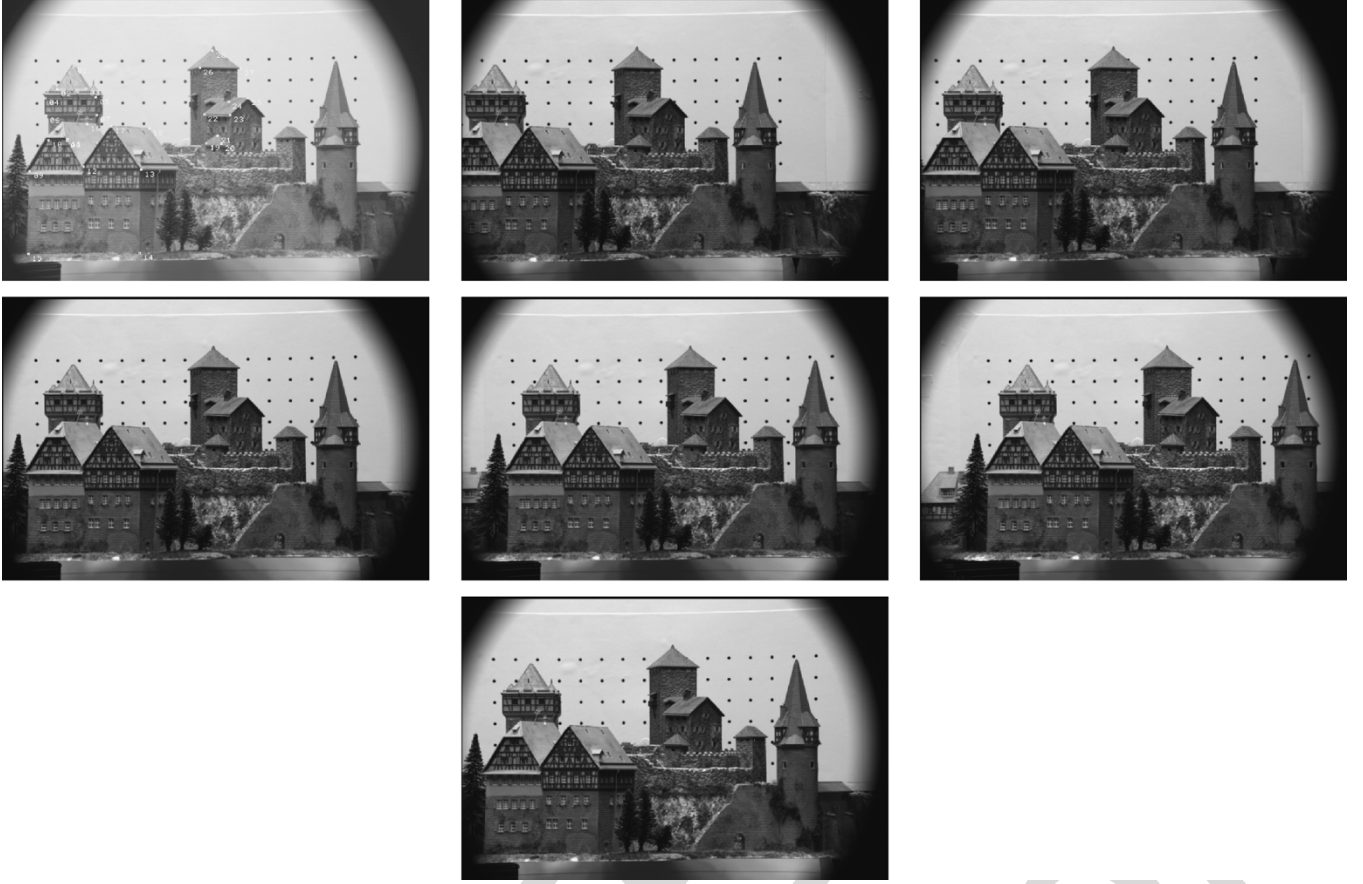


Fig. 11. Series of real castle images. The first one is the 3-D model. The others are the projective castle1–castle6 images captured at different positions.

is properly defined within the interval $[1.0, 2.0]$, good registration results can be obtained.

3. Free form surfaces are represented as sets of sparse points. All other geometric representations of free form surfaces such as triangular meshes, planar patches, or implicit or explicit representations can all be transformed to sets of points [1]. The sizes of the number of points depend on the representation accuracy required to approximate the free form surfaces. In practice, a limited number of salient points [11], [45] can be selected and used for efficient registration, as is the case for the experiments reported in this paper.
4. Let n_1 and n_2 represent the number of points in the 3-D model and in the projective image respectively. For the sake of conciseness, the ICPF algorithm employed the brute force method for the closest point search. Therefore, it retains the computational complexity, $O(n_1 n_2)$, of the traditional ICP algorithm [1]. When n_1 and n_2 are large, the algorithm is dominated in the sense of computation by the search of closest points.

B. Conclusions

In this paper, we have made three contributions. First, we have applied the rigid motion constraints to define the qualities of the traditional ICP criterion established possible point matches

TABLE IV
ACTUALLY CALIBRATED CAMERA POSITION $\hat{\mathbf{t}}$ AND THE RELATIVE CALIBRATION ERROR e_t IN PERCENTAGE OF THE CAMERA POSITION $\hat{\mathbf{t}}$ FOR THE FLS AND ICPF ALGORITHMS USING REAL IMAGES

Image	Method	$\hat{\mathbf{t}}$	$e_t(\%)$
Castle	FLS	-558.35	2.85
		-521.01	
		1712.99	
	ICPF	-558.18 -521.19 1709.15	3.05
Wall and Tower	FLS	-557.720	0.20
		-520.27	
		1988.70	
	ICPF	-557.56 -520.20 1992.70	0.24
Planar Texture	FLS	-559.07	1.08
		-522.39	
		1748.74	
	ICPF	-559.03 -523.30 1744.75	1.28

based on which different point matches can be compared and relatively good point matches are selected for motion parameter re-estimation in the projective registration of a 3-D model and its projective image; secondly, we have theoretically proved that for the 3-D–2-D point matches established by the traditional

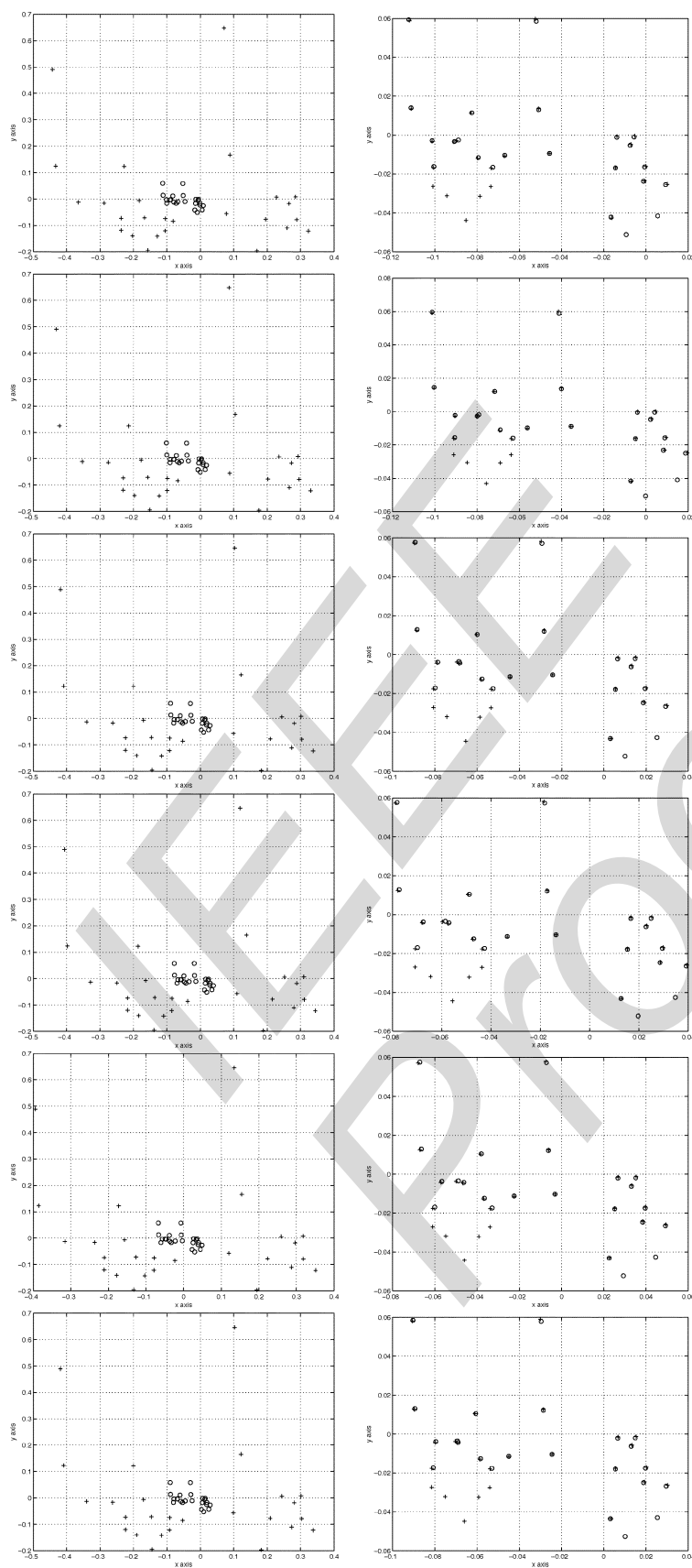


Fig. 12. Projective registration results based on real castle images. Left column: Result at iteration 1. Right column: Result after registration. From top row to bottom row: castle1–castle6.

TABLE V
ACTUALLY CALIBRATED CAMERA POSITION $\hat{\mathbf{t}}$ AND THE RELATIVE
CALIBRATION ERROR e_t IN PERCENTAGE OF THE CAMERA POSITION $\hat{\mathbf{t}}$
FOR THE FLS AND ICPF ALGORITHMS USING REAL IMAGES

Image	Method	$\hat{\mathbf{t}}$	$e_t(\%)$
Castle1	FLS	-558.35	2.85
		-521.01	
		1712.99	
	ICPF	-558.18 -521.19 1709.15	3.05
Castle2	FLS	-559.52	2.43
		-519.93	
		1721.80	
	ICPF	-539.51 -520.18 1719.12	2.57
Castle3	FLS	-518.35	2.10
		-523.15	
		1727.05	
	ICPF	-518.88 -523.09 1723.69	2.27
Castle4	FLS	-498.83	1.89
		-522.99	
		1731.76	
	ICPF	-499.39 -523.03 1729.20	2.02
Castle5	FLS	-478.99	1.75
		-522.95	
		1733.56	
	ICPF	-479.66 -523.01 1731.45	1.85
Castle6	FLS	-518.58	2.13
		-523.01	
		1707.08	
	ICPF	-518.86 -522.97 1703.78	2.29

TABLE VI
RELATIVE CALIBRATION ERROR e_t IN PERCENTAGE OF THE CAMERA
POSITION $\hat{\mathbf{t}}$ FOR THE ICPF ALGORITHM USING THE REAL CASTLE AND
CASTLE1 IMAGES WITH DIFFERENT PERCENTAGES OF APPEARING AND
DISAPPEARING POINTS. COLUMN CORRESPONDS TO THE PERCENTAGE
OF DISAPPEARING 3-D MODEL POINTS. ROW CORRESPONDS TO
THE PERCENTAGE OF APPEARING PROJECTIVE IMAGE POINTS

$e_t(\%)$	0%	5%	10%	15%	20%
0%	2.87	2.69	2.80	2.61	2.85
5%	2.95	2.91	2.76	2.71	2.81
10%	3.01	2.91	2.93	2.98	3.05
15%	3.35	3.21	3.12	3.40	3.24
20%	11.55	3.35	3.33	3.38	3.27

ICP criterion, as long as at least two of them satisfy the rigid motion constraints *exactly*, then any one of them that satisfies the rigid motion constraints *exactly* represents a real 3-D–2-D correspondence. This is in contrast with the feature matching based methods [14], [40], [42] that often take for granted that the finally established correspondences are real. Our study has explicitly revealed the conditions under which accurate 3-D–2-D

TABLE VII
RELATIVE CALIBRATION ERROR e_t IN PERCENTAGE OF THE CAMERA
POSITION $\hat{\mathbf{t}}$ FOR THE ICPF ALGORITHM USING THE REAL WALL AND TOWER
IMAGES WITH DIFFERENT PERCENTAGES OF APPEARING AND DISAPPEARING
POINTS. COLUMN CORRESPONDS TO THE PERCENTAGE OF DISAPPEARING
3-D MODEL POINTS. ROW CORRESPONDS TO THE PERCENTAGE
OF APPEARING PROJECTIVE IMAGE POINTS

$e_t(\%)$	0%	5%	10%	15%	20%
0%	0.35	0.22	0.23	0.20	0.33
5%	0.69	0.20	0.29	0.20	0.36
10%	0.84	0.88	0.30	0.66	0.44
15%	0.20	0.70	0.54	0.70	1.24
20%	38.16	36.24	33.09	1.02	0.96

correspondences can be established; and thirdly, we have experimentally validated based on both synthetic points data and real images that rigid motion constraints often render the false matches to be successfully eliminated, thus, achieving encouraging projective registration results.

However, in practice, the rigid motion constraints are often violated. The extent of violation often depends on the distribution of points and noise level corrupting data points. Even so, correct point matches can still be determined. Inappropriate setting of the parameter κ in the proposed algorithm sometimes does lead our algorithm to get stuck at a local minimum, yielding inaccurate projective registration results. However, the rigid motion constraints, the ICP criterion and the statistical model in our proposed algorithm ensure that our proposed algorithm rarely degenerates catastrophically and is capable of providing a good initialization for other even more accurate projective registration algorithms.

Since the proposed algorithm just applies the information provided by the ICP criterion for possible point match evaluation without any feature extraction from images [14], [21]–[23], [40], [42] or motion estimation from the possible point matches [30], [34] before they are evaluated, it has an advantage of easy implementation. What is more important is that evolving rigid motion constraints can ensure that the finally established 3-D–2-D point matches at least *approximately* satisfy the motion (1) and, thus, it is guaranteed that the point matches established by the ICP criterion and evaluated by the rigid motion constraints represent *somewhat* real correspondences. For these reasons, our method may open a novel avenue for the projective registration of free form surfaces, represented as sets of sparse 3-D and 2-D points. Further research is being carried out in characterizing the distribution of points for dynamically setting κ , addressing nonpure translational motion registration and improving the computational efficiency of the algorithm based on techniques such as k-D tree [9], [44] and results will be reported in the future.

ACKNOWLEDGMENT

The author would like to thank Dr. L. Thomas, Dr. E. Sherratt, and Dr. S. Garrett at The University of Wales, Aberystwyth, for their proofreading that has improved the readability of the paper. He would also like to thank the four anonymous reviewers for their constructive comments that have improved the quality of

TABLE VIII
RELATIVE CALIBRATION ERROR e_t IN PERCENTAGE OF THE CAMERA POSITION \hat{t} FOR THE ICPF
ALGORITHM USING DIFFERENT INITIALIZATIONS OF THE z COORDINATE OF THE CAMERA POSITION

	castle1	castle2	castle3	castle4	castle5	castle6	WallTower	PlanarTexture
$e_t(\%)$	3.03	2.25	2.27	2.00	1.81	2.29	0.80	1.28

the paper. Real-image data for this research were partially provided by the Calibrated Imaging Laboratory, Carnegie Mellon University.

REFERENCES

- [1] P. J. Besl and N. D. McKay, "A method for registration of 3D shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 239–256, Feb. 1992.
- [2] A. P. Blicher, S. Roy, and P. S. Penew, "LightSphere: fast lighting compensation for matching a 2D image and a 3D model," presented at the *ICPR*, vol. 2, 2004.
- [3] Y. Chen and G. Medioni, "Object modeling by registration of multiple range images," *IVC*, vol. 10, pp. 145–155, 1992.
- [4] C. M. Cyr, T. B. Sebastian, and B. B. Kimia, "2D-3D registration based on shape matching," in *Proc. IEEE Workshop MMBIA*, 2000, pp. 198–203.
- [5] P. David, D. DeMenthon, R. Duraiswami, and H. Samet, "Simultaneous pose and correspondence determination using line features," in *Proc. CVPR*, vol. 2, 2003, pp. 424–431.
- [6] —, "SoftPOSIT: simultaneous pose and correspondence determination," in *Proc. ECCV*, 2002, pp. 698–714.
- [7] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc. Ser. B*, vol. 39, pp. 1–38, 1977.
- [8] M. Dhome, M. Richetin, J. T. Lapreste, and G. Rives, "Determination of the attitude of 3D objects from a single perspective view," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-11, no. 8, pp. 1265–1278, Aug., 1989.
- [9] J. Feldmar, N. Ayache, and F. Betting, "3D-2D projective registration of free-form curves and surfaces," *CVIU*, vol. 65, pp. 403–424, 1997.
- [10] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Comm. ACM*, vol. 24, pp. 381–395, 1981.
- [11] F. Fraundorfer, S. Ober, and H. Bischof, "Toward robot localization using natural, salient image," in *Proc. 9th Computer Vision Winter Workshop*, 2004, pp. 159–166.
- [12] M. Gleicher, "Projective registration with difference decomposition," in *Proc. CVPR*, 1997, pp. 331–337.
- [13] T. Gramegna, L. Venturino, G. Cicirelli, G. Attolico, and A. Duistante, "Optimization of the POSIT algorithm for indoor autonomous navigation," *Robot. Autonom. Syst.*, vol. 48, pp. 145–162, 2004.
- [14] W. E. L. Grimson, D. P. Huttenlocher, and T. D. Alter, "Recognizing 3D objects from 2D images: an error analysis," in *Proc. CVPR*, 1992, pp. 316–321.
- [15] D. Hahnel, S. Thrun, and W. Burgard, "An extension of the ICP algorithm for modeling nonrigid objects with mobile robots," in *Proc. Int. Joint Conf. Artificial Intelligence*, 2003, pp. 915–920.
- [16] J. Y. Herve, C. Duchesne, and V. Pradines, "Dynamic registration for augmented reality in telerobotics applications," in *Proc. Int. Conf. SMC*, 2000, pp. 1348–1353.
- [17] J. Jang and K. Hong, "Self-calibration of a stereo-camera by pure translational motion," in *Proc. ICIP*, 1996, pp. 297–300.
- [18] C. Jerian and R. C. Jain, "Determining motion parameters for scenes with translation and rotation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, no. 4, pp. 523–530, Apr. PAMI-1984.
- [19] A. Johnson and S. Kang, "Registration and integration of textured 3-D data," in *Proc. 3DIM*, 1997, pp. 234–241.
- [20] T. Kanade and M. Okutomi, "A stereo matching algorithm with an adaptive window: theory and experimental results," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 7, pp. 920–932, Jul. 1994.
- [21] Y. Kita, D. L. Wilson, and J. A. Noble, "Real-time registration of 3D cerebral vessels to X-ray angiograms," in *Proc. MICCAI*, 1998, pp. 1125–1133.
- [22] Y. Kita, D. L. Wilson, J. A. Noble, and N. Kita, "A quick 3D-2D registration method for a wide-range of applications," in *Proc. ICPR*, 2000, pp. 1981–1986.
- [23] Y. Kita and N. Kita, "Position and pose detection of active camera-head in a nuclear power plant," in *Proc. IROS*, 2000, pp. 1872–1879.
- [24] R. Kurazume, K. Nishino, Z. Zhang, and K. Ikeuchi, "Simultaneous 2D images and 3D geometric model registration for texture mapping utilizing reflectance attribute," in *Proc. ACCV*, 2002, pp. 99–106.
- [25] S. Lavalée and R. Szeliski, "Recovering the position and orientation of free-form objects from image contours using 3D distance maps," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 3, pp. 378–390, Mar. 1995.
- [26] C. T. D. Lin, D. B. Goldgof, and W. C. Huang, "Motion estimation from scaled orthographic projections without correspondences," *IVC*, vol. 12, pp. 95–108, 1994.
- [27] A. Liu, E. Bullitt, and S. M. Pizer, "3D/2D registration via skeletal near projective invariance in tubular objects," in *Proc. MICCAI*, 1998, pp. 952–963.
- [28] Y. Liu and M. A. Rodrigues, "An iterative algorithm for the projective registration of free-form surfaces," in *Proc. ICIP*, 2000, pp. 497–500.
- [29] Y. Liu, M. A. Rodrigues, and B. Wei, "A novel method to cope with appearing and disappearing points for the projective registration of free-form surfaces," in *Proc. ICIP*, 2001, pp. 600–603.
- [30] Y. Liu and M. A. Rodrigues, "Geometrical analysis of two sets of 3D correspondence data patterns for the registration of free-form shapes," *J. Int. Robot. Syst.*, vol. 33, pp. 409–436, 2002.
- [31] Y. Liu and Y. Wang, "Evaluating 3D-2D correspondences for accurate camera pose estimation from a single image," in *Proc. Int. Conf. SMC*, 2003, pp. 703–708.
- [32] G. P. Penney, J. Weese, J. A. Little, P. Desmedt, D. L. G. Hill, and D. J. Hawkes, "A comparison of similarity measures for use in 2-D-3-D medical image registration," *IEEE Trans. Med. Imag.*, vol. 17, pp. 586–595, 1998.
- [33] A. Roche, G. Malandain, N. Ayache, and X. Pennec, "Multimodal image registration by maximization of the correlation ratio," INRIA, France, Tech. Rep. 3378, 1998.
- [34] M. A. Rodrigues and Y. Liu, "On the representation of rigid body transformations for accurate registration of free form shapes," *Robot. Autonom. Syst.*, vol. 39, pp. 37–52, 2002.
- [35] G. C. Sharp, S. W. Lee, and W. K. Wehe, "ICP registration using invariant features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 1, pp. 90–112, Jan. 2002.
- [36] P. W. Smith and M. D. Elstrom, "Stereo-based registration of range and projective imagery for data fusion and visualization," *Opt. Eng.*, vol. 40, pp. 352–361, 2001.
- [37] H. Surmann, A. Nuchter, and J. Hertzberg, "An autonomous mobile robot with a 3D laser range finder for 3D exploration and digitalization of indoor environment," *Robot. Autonom. Syst.*, vol. 45, pp. 181–198, 2003.
- [38] R. Y. Tsai and T. S. Huang, "Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, no. 1, pp. 13–27, Jan. PAMI-1984.
- [39] C.-L. Tsai, A. Majerovics, C. V. Stewart, and B. Roysam, "Disease-oriented evaluation of dual-bootstrap retinal image registration," in *Proc. MICCAI*, 2003, pp. 754–761.
- [40] W. Wells, "Statistical approaches to feature-based object recognition," *IJCV*, vol. 21, pp. 63–98, 1997.
- [41] R. T. Whitaker and J. Gregor, "A maximum-likelihood surface estimator for dense range data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 9, pp. 1372–1387, Sep. 2002.
- [42] P. Wlczek, A. Maccato, and R. J. P. deFigueiredo, "Pose estimation of three-dimensional objects from single camera images," *Digit. Signal Process.*, vol. 5, pp. 176–183, 1995.
- [43] P. Wunsch and G. Hirzinger, "Registration of CAD-models to images by iterative inverse perspective matching," in *Proc. ICPR*, 1996, pp. 78–83.
- [44] Z. Zhang, "Iterative point matching for registration of free-form curves," INRIA, France, Tech. Rep. 1658, 1992.
- [45] M. Zuliani, C. Kenney, and B. S. Manjunath, "A mathematical comparison of point detectors," in *Proc. Conf. CVPR Workshop: Image and Video Registration (IVR)*, 2004, p. 172.



Yonghuai Liu received the Ph.D. degree from Northwestern Polytechnical University, [AUTHOR: PLEASE PROVIDE CITY] China , and the Ph.D. degree from The University of Hull, Hull, U.K., in 1998 and 2001, respectively.

He joined the University of Wales, Aberystwyth, U.K., in September 2001 as a Lecturer. From April to September 1999, he was a Research Assistant at The University of Hull. From March 2000 to September 2001, he was a Research Fellow at Sheffield Hallam University [AUTHOR: PLEASE

PROVIDE CITY/COUNTRY]. He has published more than 70 papers in international conference proceedings and journals. He was the Guest Editor for the 2002 special issue of the *Computer Vision and Image Understanding Journal* on the registration and fusion of range images. His primary research interests lie in machine vision, image registration, motion estimation, pattern recognition, image processing, artificial intelligence, and intelligent systems.

IEEE
Proof