# A Unified Approach Combining Photometric and Geometric Information for Pose Estimation

Pierre Georgel, Selim Benhimane, Nassir Navab
Computer Aided Medical Procedures & Augmented Reality (CAMP),
Technical University of Munich (TUM),
Boltzmannstrasse 3, 85748 Garching, Germany
{georgel,benhiman,navab}@in.tum.de

**Abstract**

In this paper, we present a novel approach for the relative pose estimation problem from point correspondences extracted from image pairs. Unlike classical algorithms, such as the Gold Standard algorithm, the proposed approach ensures that the matched points are photo-consistent throughout the pose estimation process. In fact, common algorithms use the photometric information to extract the feature points and to establish the 2D point correspondences. Then, they focus on minimizing, in a non-linear scheme, geometric distances between the projection of reconstructed 3D points and the coordinates of the extracted image points without taking the photometric information into account. The approach we propose in this paper merges geometric and photometric information in a unified cost function for the final non-linear minimization. This allows us to achieve results with higher precision and also with higher convergence frequency. Extensive experiments with ground truth on synthetic data show the superiority of the proposed approach in terms of robustness and precision. The simulation results have been confirmed by several tests on real image data.

## 1 Introduction

In Computer Vision, relative pose estimation corresponds to the task of finding the geometric transformation between two cameras. Using two images each acquired by one camera, it is possible to use a set of corresponding feature points to estimate the pose parameters, i.e. the relative rotation and translation between the two cameras. This task is a core problem of several computer vision applications (such as 3D reconstruction, vision-based control or augmented reality) and it has been extensively studied for the last decades. Since the seminal work of Longuet-Higgins [9] where an 8-point algorithm was proposed to compute the pose via the essential matrix, many works have been published either to generalize it to the non-calibrated case [4, 1], or to improve the robustness [6, 12] or as proposed recently, to solve it efficiently in a closed-form algorithm with the minimal set of five points [11]. The standard scheme that one can find in reference Computer Vision books [7, 2] or in earlier work (e.g. [8]) is to actually use several corresponding points (from some tens to few hundreds) and to minimize a least-squares cost function making use of all available correspondences. In practice, this allows the pose estimation

BMVC 2008 doi:10.5244/C.22.14

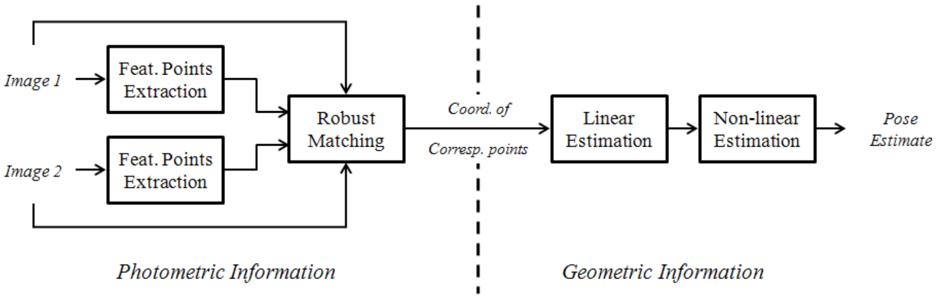

Figure 1: Classical approaches set aside the image photometric information once the matching has been established. Only geometric information is used in the pose estimation.

as well as the 3D points reconstruction to be robust and precise. A data normalization precedes a linear estimation or a closed-form solution which generally serves as initialization of such non-linear minimization. The most commonly used algorithm to perform this task is the well known Gold Standard algorithm [7]. Based on minimizing the reprojection error, this algorithm allows to obtain the Maximum Likelihood estimate of the fundamental matrix (and can be easily adapted to the essential matrix) whose decomposition provides an estimation of the pose. As depicted by figure 1, this algorithm makes only use of the coordinates of the image points. No photometric information (color, texture, image gradients,...) is used once the matching of the feature points has been established.

The Gold Standard algorithm works well in practice and gives satisfactory results in most cases. Recent improvements have been proposed such that the linear estimation has a better conditioned measurement matrix [12] and such that the non-linear optimization quickly and efficiently converges toward the global minimum [5]. In the presence of small number of noisy feature points, even if a robust estimation [3] has been applied to remove outliers in the matching process, it is hard to recover a precise pose. It is possible to get a geometrically correct and globally optimal pose, but it can be, in some cases, far from the real one. This is mainly due to the fact that such approaches do not guarantee a photometric consistency with respect to the images.

In this paper, we introduce an additional constraint to the traditional reprojection error during the final non-linear optimization. To state it simply: the points will be additionally constrained to have the same appearance in both images.

## 2 Theoretical Background

**Notations:** Bold upper-case variables are matrices. Bold lower-case variables are vectors. $\mathscr{X}$ are 3D points in homogeneous coordinates. $\mathbf{w}$ is the projection function transforming a 3D homogeneous point to a 2D homogeneous point. $\mathscr{I}$ is the image function that takes a 2D point and gives an intensity. We suppose that the reference frame of the 3D points $\mathscr{X}$ is aligned with the first camera that pictured $\mathscr{I}_1$. Its projection matrix is then the identity matrix $\mathbf{I}_{4\times4}$. The transformation matrix between the first and the second camera that pictured $\mathscr{I}_2$ is $\mathbf{T} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \in \mathbb{SE}(3)$ where $\mathbf{R} \in \mathbb{SO}(3)$ and $\mathbf{t} \in \mathbb{R}^3$ are the rotation matrix and the translation vector.

We are interested in estimating the rigid motion between two calibrated views using a set of 2D points extracted from the images. The relation between two corresponding points $\mathbf{p}$ (in the first image) and $\mathbf{p}'$ (in the second image) can be described using the essential matrix $\mathbf{E} = [\mathbf{t}]_{\times} \mathbf{R}$:

$$\mathbf{p}'^{\top}\mathbf{K}'^{-\top}\mathbf{E}\mathbf{K}^{-1}\mathbf{p} = 0\,, \tag{1}$$

where $\mathbf{K}$ and $\mathbf{K}'$ are the camera intrinsic parameter matrices of the first and the second camera. The estimation of this matrix provides (via a simple decomposition) the translation $\mathbf{t}$ (up to a scale) and the rotation $\mathbf{R}$. The essential matrix is generally computed using the following steps. Given a set of putative matching points (e.g. obtained using [10]), it is possible to remove outliers using a robust estimation [3] based on estimating the essential matrix with a minimal sets of points (e.g. [11]). Then, a linear estimation is performed (e.g. using [6] or [12]) in order to have an initial estimate of the pose ($\mathbf{R}$, $\mathbf{t}$) and an initial reconstruction of the 3D points (up to a scale) for the final non-linear minimization.

The non-linear estimation iteratively updates the motion parameters and the reconstructed 3D points by computing increments that reduce the distance between projection of the 3D points and their corresponding image points. The sum-of-squared differences cost function generally used is the following:

$$\arg\min_{\mathscr{X}_i,\mathbf{T}} \quad \sum_i d\left(\mathbf{p}_i, \mathbf{K}\mathbf{w}\left(\mathscr{X}_i\right)\right)^2 + d\left(\mathbf{p}'_i, \mathbf{K}'\mathbf{w}\left(\mathbf{T}\mathscr{X}_i\right)\right)^2\,, \tag{2}$$

This error minimization only ensures geometric validity of the structure consistency of the points. It is important to note that in this framework the result of the feature points detection is never corrected based on photometric information.

# 3 Proposed Method

Due to some factors that generally provide inaccurate feature points localization (such as inappropriate threshold setting in the robust outliers removal, motion blur or noise in the images), the standard way that we described above lacks precision. This is done to the fact that the feature point positions are only corrected during the non-linear minimization via the optimization of the 3D point positions. This correction is only based on geometric constraints related exclusively on the points coordinates. Therefore, a large amount of information is set aside since no photometric information is used once the matching has been established.

To make use of all available information throughout the pose estimation process, we propose to alter the cost function used in the non-linear minimization such that photometric information is also taken into account. Here We describe how to incorporate this function in the final pose refinement step. Figure 2 gives an overview of our method.

## 3.1 Combining Geometric Distances with Intensity Differences

In order to incorporate photometric information, we consider the fact that the neighborhood of the 2D points $\mathbf{p}_i$ and $\mathbf{p}'_i$ should be photometrically consistent. To enforce their consistency, we optimize the pose parameters and the 3D points such that, both the cost

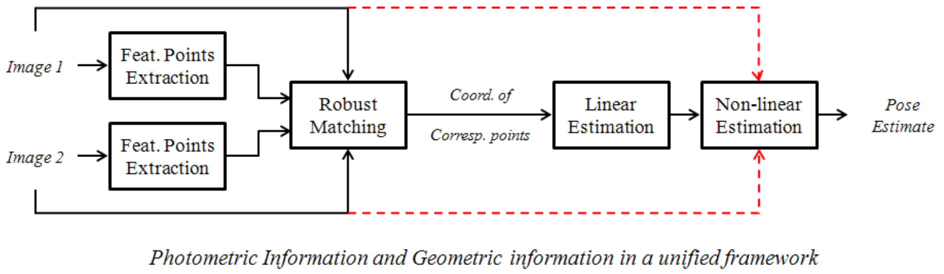

Figure 2: Our approach uses the photometric information not only for the matching but also for the non-linear estimation where photometric and geometric cues are combined.

function defined in equation (2) and the sum-of-squared differences of the intensities of the projected neighboring points are minimized. Here, some issues should be carefully taken into account:

1. How can we define the neighborhood of the 2D points in the two images ?
2. At which stage the photometric term should be used ?
3. How should the geometric and the photometric terms be weighted ?

Concerning the first issue, we define the neighborhood of the feature points in the two images using samples on the tangent plane of the reconstructed 3D points since any surface can be locally approximated as planar, as shown in Figure (3). We will show later that even a fronto-parallel approximation of the tangent planes is enough to improve the results. As a consequence the neighborhood of 2D feature points will be adapted when the pose and the 3D point locations are refined during the non-linear minimization. If we denote by $\mathscr{Y}_{ij}$ a point of the neighborhood $\Omega_i$ of $\mathscr{X}_i$ ($\Omega_i$ is represented as the tangent plane to $\mathscr{X}_i$ with $\mathbf{n}_i$ its normal). The following term should be minimized:

$$\sum_{j\in\Omega_i}\left(\mathscr{I}_1\left(\mathbf{Kw}\left(\mathscr{Y}_{ij}\right)\right)-\mathscr{I}_2\left(\mathbf{K}'\mathbf{w}\left(\mathbf{T}\mathscr{Y}_{ij}\right)\right)\right)^2 \tag{3}$$

In general, the non-linear minimization is based on a first-order Taylor expansion of a cost function (such as Gauss-Newton or Levenberg-Marquardt). Since the image function has been introduced, the convexity assumption (needed by such minimization methods in order to succeed) is very local for the term (3). The image patches obtained by projecting the 3D neighborhood in the two images should project to the same physical object in order to have a locally convex cost function. When using the pose resulting from the linear estimation and performing the 3D triangulation form the image points $(\mathbf{p}_i, \mathbf{p}_i')$, the distances between measured points and their reprojected 3D points $\mathscr{X}_i$ can be too large (depending on the inaccuracy of the point extraction). The table 1, which was obtained using simulated data, confirms this argument. In presence of a bad initialization, the projection of the neighbors in each image might give two patches that will not overlap the same area of the observed scene. Consequently, for the second issue, we added a test based on the geometric distance and based on the Normalized-Cross Correlation (NCC) between the

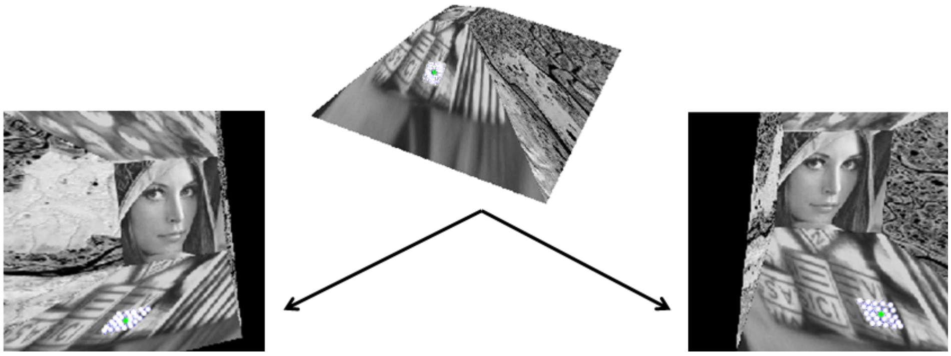

Figure 3: The Neighbors $\mathscr{Y}_{ij}$ (in white) defined in 3D around triangulated point $\mathscr{X}_i$ (in green), they are projected in the image to create $\texttt{Patch}_\texttt{i}$ and $\texttt{Patch}'_\texttt{i}$ as shown in left and right pictures.

patches in order to ensure that the projected patches are close enough. Let us denote by $\texttt{Patch}_\texttt{i} = \left\{\mathscr{I}_1\left(\mathbf{Kw}\left(\mathscr{Y}_{ij}\right)\right), j \in [1..m]\right\}$ and by $\texttt{Patch}'_\texttt{i} = \left\{\mathscr{I}_2\left(\mathbf{Kw}\left(\mathbf{T}\mathscr{Y}_{ij}\right)\right), j \in [1..m]\right\}$, with $m$ the number of samples, the ordered sets of intensities obtained by projecting the 3D point $\mathscr{X}_i$ in the first and in the second image respectively. The image information of the corresponding points $\mathbf{p}_i$ and $\mathbf{p}'_i$ will be considered when $\delta_i = 1$ and will not be considered when $\delta_i = 0$ where:

$$\delta_i = \begin{cases} 1 & \text{if} \quad \& \begin{cases} d\left(\mathbf{p}_i, \mathbf{Kw}\left(\mathscr{X}_i\right)\right) < \tau_1 \ \&\ d\left(\mathbf{p}'_i, \mathbf{K}'\mathbf{w}\left(\mathbf{T}\mathscr{X}_i\right)\right) < \tau_1 \\ \mathrm{NCC}\left(\texttt{Patch}_\texttt{i}, \texttt{Patch}'_\texttt{i}\right) > \tau_2 \end{cases} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

| #Points/Noise | 0.01 | 0.1 | 1.0 | 2.0 |
|---|---|---|---|---|
| 8 | 0.3619 | 3.9054 | 96.3384 | 179.4612 |
| 10 | 0.2066 | 1.9684 | 17.0605 | 24.6839 |
| 25 | 0.1002 | 1.0511 | 10.4050 | 18.5587 |
| 50 | 0.0660 | 0.6916 | 6.0568 | 14.0087 |

Table 1: Evolution of the mean residual error w.r.t. ground truth (in pixels) over 200 runs after applying the 8-points algorithm and the optimal triangulation, w.r.t. the number of points and the Gaussian noise.

Finally, for the last issue, since the geometric and photometric data are heterogeneous, they should be carefully integrated in a unified cost function. If we simply stack them together, we will have a massive scale differences. In fact, intensity differences could vary between $[-255, 255]$ (when the pixel intensity is coded in 8 bits) while geometric distances are expressed in pixels. In order to obtain a more uniform observation, we scale the geometric distance by the inverse of the variance of the vector $\left[\left(\mathbf{Kw}\left(\mathscr{X}_i\right) - \mathbf{p}_i\right)^\top \left(\mathbf{K}'\mathbf{w}\left(\mathbf{T}\mathscr{X}_i\right) - \mathbf{p}'_i\right)^\top\right]^\top$. We do the same for the photometric distance with the vector $\left[\left(\mathscr{I}_1\left(\mathbf{Kw}\left(\mathscr{Y}_{ij}\right)\right) - \mathscr{I}_2\left(\mathbf{K}'\mathbf{w}\left(\mathbf{T}\mathscr{Y}_{ij}\right)\right)\right)^\top\right]^\top$. These scales are computed at initialization. This will compensate not only for the difference of scale but also for the difference of size (we have much more pixels than points coordinates). The scaling factors $\alpha_g$ for the geometric term and $\alpha_p$ for the photometric term will then be included in the unified cost function.

## 3.2 Unified Cost Function

We take into account the issues explained above and we modify the cost function of equation (2) to the following form:

$$\arg\min_{\mathscr{X}_i,\mathbf{T}} \quad \alpha_g \mathscr{D}_g\left(\mathscr{X}_i,\mathbf{T}\right) + \alpha_p \mathscr{D}_p\left(\mathscr{X}_i,\mathbf{T}\right) . \qquad (5)$$

In addition to the traditional geometric error term :

$$\mathscr{D}_g = \sum_i d\left(\mathbf{p}_i, \mathbf{Kw}\left(\mathscr{X}_i\right)\right)^2 + d\left(\mathbf{p}'_i, \mathbf{K}'\mathbf{w}\left(\mathbf{T}\mathscr{X}_i\right)\right)^2 , \qquad (6)$$

a photometric error term has been added:

$$\mathscr{D}_p = \sum_i \delta_i \sum_j \left(\mathscr{I}_1\left(\mathbf{Kw}\left(\mathscr{Y}_{ij}\right)\right) - \mathscr{I}_2\left(\mathbf{K}'\mathbf{w}\left(\mathbf{T}\mathscr{Y}_{ij}\right)\right)\right)^2 . \qquad (7)$$

In the next section, we provide some details of the implementation. The parameters used in the simulation are presented in the experimental results section.

## 3.3 Implementation Details

The neighborhoods are defined as a regular grid around each 3D point $\mathscr{X}_i$ oriented according to the given normal $\mathbf{n}_i$. Each neighborhood has a specific size, the edge length $s_i$ is calculated based upon an upper bound $p$ of the edge length of a patch in the image, as shown in algorithm (1) . The number of elements in the neighborhood can be adjusted depending on the desired sampling.

Since the area of the neighborhood will not be updated during the minimization, the variation of scale within the 3D structure must be limited in order to insure that a patch does

**Algorithm 1** Compute the wanted edge length $s_i$ of a 3D neighborhood, given a 3D points $\mathscr{X}_i$, its normal $\mathbf{n}_i$, a pose $\mathbf{T}$ and the maximum size of the patch $d$ in the image in pixels

```
Require: $\mathscr{X}_i$ ,$\mathbf{n}_i$, $\mathbf{T}$, $d$
  $s_i \leftarrow 1$; $max_{border} \leftarrow d$
  repeat
    $s_i \leftarrow s_i * d / max_{border}$
    Create a neighborhood of size $s$ in 3D of $\mathscr{X}_i$ and $\mathbf{n}_i$
    Create Patch, Patch' by projecting the neighborhood in both image
    Compute $size_{border}$ the distance between each corners of Patch
    Compute $size'_{border}$ the distance between each corners of Patch'
    $max_{border} \leftarrow \max\left(size_{border}, size'_{border}\right)$
  until $|max_{border} - d| > \varepsilon$
```

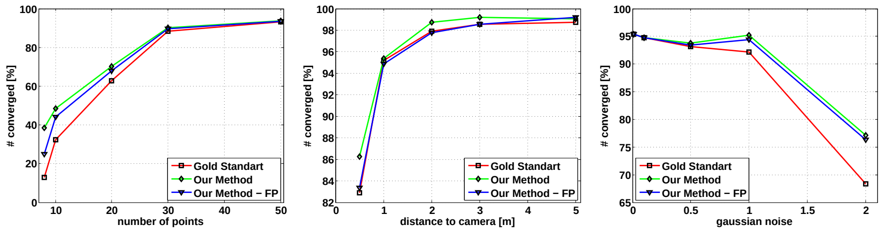

Figure 4: Convergence rate in percent; the first graph shows results with an increasing number of points and a Gaussian noise of 0.5; the second graph shows the effect of an increasing distance from the scene with 50 points and a Gaussian noise of 0.5; the last graph shows the results with 50 points and varying Gaussian noise from 0.01 to 2; it shows that the additional intensity information enables our approach to be more robust even with a fronto-parallel assumption (FP).

not become too small or too large within the image. This could happen if the structure's scale shrinks which would lead to an increase in the patch's relative size in the image. In order to prevent such behaviors, we force the norm of translation $\mathbf{t}$ to be always equal to 1. During the optimization, $\mathbf{T}$ and the 3D points $\mathscr{X}_i$ are modified with the increments $\Delta\mathbf{T}$ and $\Delta\mathscr{X}_i$ given by the Gauss-Newton as follow:

$$\left\{ \begin{array}{ccc} & \mathbf{T} & \leftarrow & \Delta\mathbf{T}\,\mathbf{T} \\ \forall i, & \mathscr{X}_i & \leftarrow & \frac{\widehat{\mathscr{X}_i}+\Delta\mathscr{X}_i}{\|\mathbf{t}\|} \\ & \mathbf{t} & \leftarrow & \frac{\mathbf{t}}{\|\mathbf{t}\|} \end{array} \right. \tag{8}$$

This update will not modify the value of (6) because the geometric cost function is independent of the scale.

# 4 Experiments and Results

The implementation of our method have been performed using Matlab. The algorithm is a classic gauss-newton which stops after 50 iterations or when the error evolves less than $10^{-9}$. For initialization we used the normalized 8-point to get the pose and we obtain the 3D points using the optimal triangulation. For all the experiments, we used a maximum patch edge length of 35 pixels and a neighborhood resolution ($m$) of $35 \times 35$. For the threshold introduced in (4) we used $\tau_1 = 15$, such a threshold will not make our method dependent on the bad behavior of the Gold Standard since we can easily assume that it brings the points $\mathscr{X}_i$ in such a threshold, and $\tau_2 = 0.3$ which is there only to guarantee that the two patches are close enough to each other. We first discuss the experiments on synthetic data and then on real images.

## 4.1 Synthetic Experiments

In order to generate the synthetic images we used a 3D pyramid with the top edge cut out to create a flat area, the object used is pictured in Figure 3. All the faces were textured

| Algo./Points | 8 | 10 | 20 | 50 |
|---|---|---|---|---|
| Gold Standard | 0.3804 -5.8% | 0.3515-2.6% | 0.3137-2.6% | 0.2281-1.7% |
| Our Method | 0.2713-94.2% | 0.2830-97.4% | 0.2823-97.4% | 0.2191-98.3% |
| Gold Standard | 0.3764-14.5% | 0.3515-5.6% | 0.3133-2.6% | 0.2281-3.3% |
| Our Method-FP | 0.3283-85.5% | 0.3122-94.4% | 0.2944-97.4% | 0.2248-96.7% |

| Algo./Noise | 0.01 | 0.5 | 1.0 | 2.0 |
|---|---|---|---|---|
| Gold Standard | 0.0045-40.6% | 0.2281-1.7% | 0.4572- 1.7% | 0.9688- 4.5% |
| Our Method | 0.0045-59.4% | 0.2191-98.3% | 0.4010-98.3% | 0.5998- 95.5% |
| Gold Standard | 0.0045-47.5% | 0.2281-3.3% | 0.4571-2.3% | 0.9714-6.0% |
| Our Method-FP | 0.0045-52.5% | 0.2248-96.7% | 0.4185-97.3% | 0.6631-94.0% |

Table 2: Upper table: Comparison of the Gold Standard against our method, in presence of a Gaussian noise (0.5) with increasing point pairs; Lower table: Comparison of the Gold Standard against our method with 50 points in presence of a varying Gaussian noise; the first value represents the mean residual w.r.t. ground truth, the percentage displays the proportion by which a method out performs the other. It shows precision improvements and repeatable performance with or without the FP assumption.

using real images. Harris corners were selected in the first image and transfered to the second image using the true transformation $\mathbf{T}$. This creates a set of true correspondences with enough textured points. For each experiments we tested our method using the exact normals and also using a fronto-parallel assumption (i.e. $\mathbf{n} = [0,0,1]^\top$). We compare our result to the one obtained with the Gold Standard.

We consider that an approach has converged when the resulting pose $\mathbf{T}$ reprojects the exact 3D Points $\mathscr{X}$ in a range inferior to $\sigma$ (the variance of the Gaussian noise) of correct image points $\mathbf{p}'$. For example with a gaussian noise of 0.5, if the residual is inferior to 0.5, it has converged. The numerical results of two different methods, as in table (2), are compared only when the two approaches have converged.

The first experiment corresponds to 625 different poses with large variation of parameters, using a Gaussian noise's variance of 0.5. Figure 4(a) sustains that the additional information represented by the difference of intensity avails a better convergence rate, especially with a low number of points. Table 2 demonstrates that our approach performs better and improves the precision with or without known normals.

We then test the behavior of our algorithm against increasing noise in the extracted features. For each level of noise, we use 625 different poses and 50 points correspondences. The result are summarized in figure 4(c) and table 2. When the noise level is small our approach has the same convergence rate and precision as the Gold Standard . When the noise increases the Gold Standard convergence rate falls faster than with our methods.

The next batch of experiments targeted the stability with respect to the distance to the scene. We used 50 points, a Gaussian noise of 0.5, and 625 poses for each depths. The result in figure 4(b) shows again that our approach performs better than the Gold Standard even with the fronto-parallel assumption.

Finally, experiments with presence of blur and noise in the image were conducted. It has shown that the convergence rate of our approach was only affected when there were

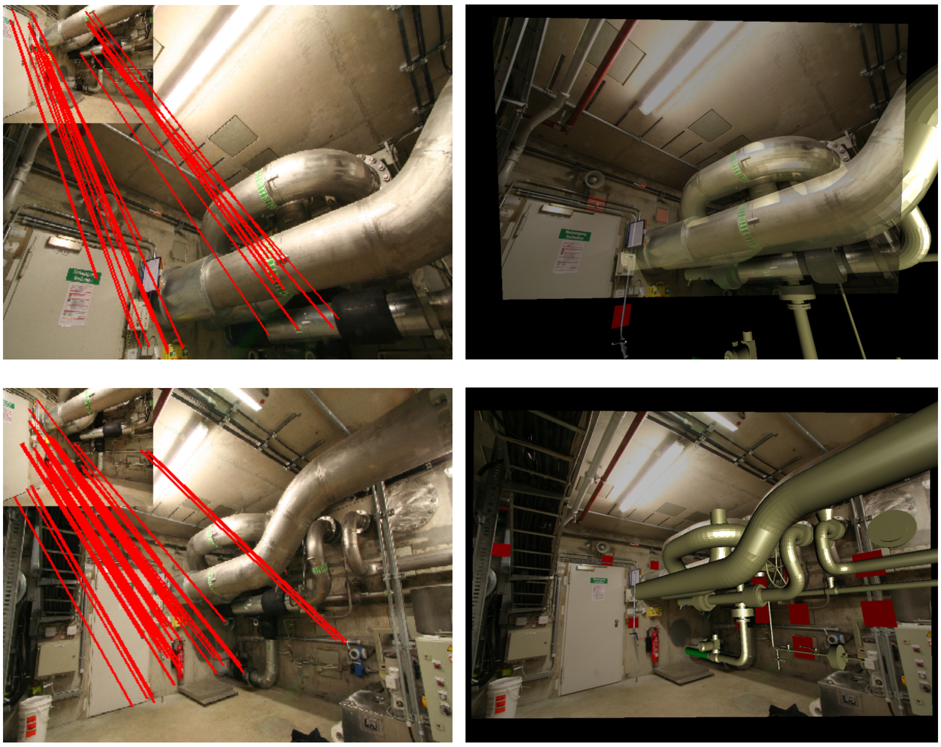

Figure 5: Real experiments: the left hand graphics represents the corresponding pair of points corrected using our method, the upper snapshot is the original keyframe; the right hand images display the resulting augmentation displayed in a VRML viewer.

massive perturbations in the image intensities. The accuracy was naturally degraded by the noise and blur. It should be noted that in the fronto-parallel approximation's case the performances were less affected than in the case where normal were known. This can be explained by the fact that the influence of the approximations of the normal is larger than noisy or blurry measure within the image.

## 4.2 Real Scenes Experiments

We tested our approach to obtain augmented images of an industrial compound. In order to recover the 3D pose of the image, we used an image manually registered to the 3D model. Using SIFT and a robust estimator we obtained a set of correct correspondences as shown in 5. Then we applied our algorithm to these correspondences. We used the fronto-parallel assumption during the non-linear estimation. Once the rotation, and the translation are recovered the length of baseline is recovered manually. This shows one of the possible application of our algorithm. It should be pointed out that the extractor of the original SIFT (difference of Gaussian) might not give the best features for our algorithm since it does not guarantee that the extracted point is locally well textured.

## 5 Conclusion

In this paper, we showed that merging geometric and photometric information in a unified cost function at the final non-linear minimization for the relative pose estimation process constraints better the results. As shown through the experiments on synthetic data, the proposed method is valuable for anyone using the classical Gold Standard algorithm for relative pose estimation since by simply modifying the cost function based reprojection error by the one we propose, the results will be more precise. Our method has been successfully applied to register image pairs with large baseline in order to superimpose virtual augmentations.

## References

[1] O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig. In *European Conf. on Computer Vision*, pages 563–578, 1992.

[2] O. Faugeras and Q. T. Luong. *The geometry of Multiple Images*. MIT Press, Cambridge, MA, 2001.

[3] M.A. Fischler and R.C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[4] R. Hartley. Estimation of relative camera positions for uncalibrated cameras. In *European Conf. on Computer Vision*, pages 579–587, 1992.

[5] R. Hartley and F. Kahl. Global optimization through searching rotation space and optimal estimation of the essential matrix. In *IEEE Int. Conf. on Computer Vision*, 2007.

[6] R. I. Hartley. In defense of the eight-point algorithm. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(6):580–593, 1997.

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

[8] B. K. P. Horn. Relative orientation. *International Journal of Computer Vision*, (5):59–78, 1990.

[9] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. 293:133–135, September 1981.

[10] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[11] David Nistér. An efficient solution to the five-point relative pose problem. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26(6):756–777, 2004.

[12] F. C. Wu, Z. Y. Hu, and Duan F. Q. 8-point algorithm revisited: Factorized 8-point algorithm. In *IEEE Int. Conf. on Computer Vision*, 2005.