How to Augment the Second Image? Recovery of the Translation Scale in Image to Image Registration

Pierre Georgel* CAMP TU München
Pierre Schroeder CAMP TU München
Selim Benhimane CAMP TU München
Mirko Appel† Siemens CT
Nassir Navab CAMP TU München

ABSTRACT

In this paper, we present an automatic pose estimation (6 DoF) technique to augment images using keyframes pre-registered to a CAD model. State of the art techniques recover the essential matrix (5 DoF) in an automatic manner, but include a manual step to align the image with the CAD reference system because the essential matrix does not provide the scale of the translation. We propose using planar structures to recover this scale automatically and to offer immediate augmentation. These techniques have been implemented in our augmented reality software. Qualitative tests are performed in an industrial environment.

1 INTRODUCTION

Photo-based Augmented Reality (AR) is an active field in the community [6, 7, 8, 3], where still images are augmented rather than video stream. Augmenting image is sometimes the only possibility to use AR technologies, for example if someone wants to augment satellite images he will have only access to a set of images not a continuous video stream. Also for many industrial applications, having real-time on site video augmentation is an overkill. For example for plants augmentation one will need a continuous access to the 3D model which tends to be extremely large, therefore might require a high bandwidth connection. Furthermore, pictures are easy to store because of their small size; and practical to share using for example emails. In particular for tasks like discrepancy check [3, 11] where the objective is to find differences between built item and the Computer Aided Design (CAD) model, the use of still-images provides some advantages, for example the possibility to annotate them. An engineer will document/annotate an image augmented with a CAD component rather than document the 3D model or mark a video stream.

The core task of an augmented reality (AR) system is to register the virtual to the real world. This transformation has 6 degrees of freedom (DOF) and is called full pose in opposition to relative pose [4] which has 5 DOF. When the full pose between the two reference systems is computed, one can augment the image with virtual information. The use of fiducial markers [2] is a common approach for obtaining this registration [8]. Other approaches can include the use of edges [1]; this involves a perfect 3D model which is rarely applicable in the industry since available 3D model often includes discrepancies. Features points [5, 9] could also be used when textured models are available. We propose to use a single keyframe to compute the full pose of the target image, meaning that we will present automatic method to go from a relative pose to a full one.

Many AR systems are based on keyframes [10, 9], where keyframes are images registered off-line to the 3D model to be used in an application such as tracking. The first method uses a single keyframe and involves projection of feature points on the 3D model to recover the full pose. The second uses a set of keyframes in order to have access to the full pose by registering the target image to all the keyframes. Compared to previous body of work our method focuses on a limited amount of 3D information, that are judged reliable. This avails to create augmentation automatically without tempering the scene with fiducials. This is extremely important for a broader use of AR solution in the industry since the software have to be operated by non-experts. Therefore it should require only a short period of training and be as automated as possible. Additionally our method could be used to create a set of keyframes without involving a complex structure from motion algorithm. The presented techniques have been implemented in the software framework described in [3].

*e-mail:georgel@in.tum.de
†e-mail:Mirko.Appel@Siemens.com

2 THEORETICAL BACKGROUND

Using a set of image points correspondences $(\mathbf{p},\mathbf{p}')$ we can recover the Essential matrix $\mathbf{E}$ that represents the points relation between two views.The Essential matrix can be decomposed as a skew matrix that stores the translation information between the two views $\mathbf{t}$ and an orthogonal matrix $\mathbf{R}$ that represents the rotation between the two views.

$$\mathbf{E} = [\mathbf{t}]_\times \mathbf{R} \tag{1}$$

Unfortunately, this decomposition is independent of the translation scale. In order to align the second view to the 3D reference frame one will have to add some 3D information. We will note with $s>0$ the scale of the translation. So the projection from the reference coordinate system and the second view is $\mathbf{P} = \begin{bmatrix} \mathbf{R} & s\mathbf{t} \end{bmatrix}$.

The common approach to recover s is to use a known 3D distance: we compute the norm of a 3D reconstructed segment visible in both image and we use the ratio between the obtained norm and the known 3D distance. Since this distance is proportional to the norm of the translation used for the reconstruction; one can easily compute the s from the known 3D distance.

It is also trivial to input the correspondence of an image point $\mathbf{p}$ and a 3D point $\mathscr{X}$ expressed in the reference frame. Using the projection formula, one can compute the true translation by subtraction.

3 AUTOMATIC SCALE RECOVERY FROM PLANE

Planar structures such as walls are considered to be one of the most reliable components of industrial components, therefore good features to match in order to align the target image to the CAD reference frame. By triangulating the points $(\mathbf{p},\mathbf{p}')$ between the two views, dominant planar structures can be extracted. Lets π be an extracted plane from the images that correspond to a plane Π from the CAD model, with d' (resp. d) the distance of the π (resp. Π) to the first camera. Let first show that the normal $\mathbf{n}$ of the reconstructed planes π is independent of s

$$\begin{aligned}\forall \mathscr{X}_1,\mathscr{X}_2,\mathscr{X}_3 \in \Pi,\, \mathbf{n} \quad &\propto [\mathscr{X}_3-\mathscr{X}_1]_\times (\mathscr{X}_2-\mathscr{X}_1)\\ &= [z_3\mathbf{m}_3 - z_1\mathbf{m}_1]_\times (z_2\mathbf{m}_2 - z_1\mathbf{m}_1)\\ &= s^2 \left[\frac{\mathbf{a}_3^\top\mathbf{b}_3}{\|\mathbf{a}_3\|^2}\mathbf{m}_3 - \frac{\mathbf{a}_1^\top\mathbf{b}_1}{\|\mathbf{a}_1\|^2}\mathbf{m}_1\right]_\times \left(\frac{\mathbf{a}_2^\top\mathbf{b}_2}{\|\mathbf{a}_2\|^2\mathbf{m}_2} - \frac{\mathbf{a}_1^\top\mathbf{b}_1}{\|\mathbf{a}_1\|^2}\mathbf{m}_1\right).\end{aligned} \tag{2}$$

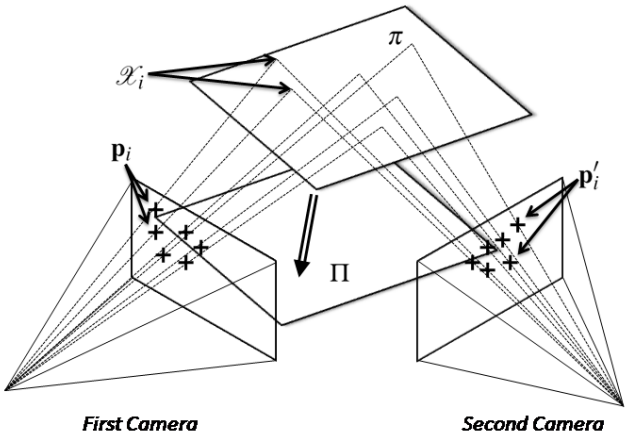

Figure 1: Scale from plane: plane π is computed triangulated points $\mathscr{X}_i$ from $(\mathbf{p}_i, \mathbf{p}'_i)$ and then is matched to Π to obtain the scale s.

So the matching between reconstructed plane and the one of the 3D model will be straight forward. Using the known distance between the camera and the plane Π we can recover the translation scale as follow:

$$\begin{array}{lll} \forall \mathscr{X} \in \Pi, \quad d & = -\mathbf{n}^\top \mathscr{X} & = -\mathbf{n}^\top (z\mathbf{m}) \\ & = -s\frac{\mathbf{n}^\top \mathbf{a}^\top \mathbf{b}}{\|\mathbf{a}\|^2}\mathbf{m} & = s\,d' \\ \Rightarrow \qquad s & = \frac{d}{d'}. & \end{array} \tag{3}$$

This relation is sumarized in figure 1.

3.1 Implementation Details

To extract the scale of the baseline using planar structure, first the pairs of points $(\mathbf{p}_i, \mathbf{p}'_i)$ are triangulated to obtain their 3D locations $\mathbf{X}_i$ using $\mathbf{R}$ and $\mathbf{t}$ (with $\|\mathbf{t}\| = 1$). Then the dominant plane is extracted from the cloud of points. This plane is segmented using RANSAC with a 3 points algorithm ($\mathbf{n} \propto \mathscr{X}_1\mathscr{X}_2 \times \mathscr{X}_1\mathscr{X}_3$, $d = -\mathbf{n}^\top \mathscr{X}_1$) and the distance to the plane ($d(\mathscr{X}, \pi) = \left|\mathbf{n}^\top \mathscr{X} + d\right|$)as a criteria to decide if a point is on the plane or not. This gives the normal $\mathbf{n}$ of the plane that is matched with the one from the 3D model using a simple dot product. Finally, we use equation (3) to compute scale of the baseline s.

4 Conclusion

In this paper, we presented a novel approach to compute the scale of the baseline extracted from an essential matrix. This problem has not been addressed within the computer vision community, because there was no interest in augmentation of the image pairs by virtual models. AR needs the right scale for correct superimposition of virtual (CAD) data. We show applications of the presented method on pictures of an industrial plant as pictured in figure 2. The software used to perform these tests is currently being experimented on a building site to fellow up the quality of built structures, bringing AR one step closer to an everyday use.

Acknowledgements

Werner Biemann, Ralf Keller, Martin Neuberger, Erwin Rusitschka and Stefan Schröter from Areva NP are thanked for their help and support; and Harald Türck from TUM for providing pictures.

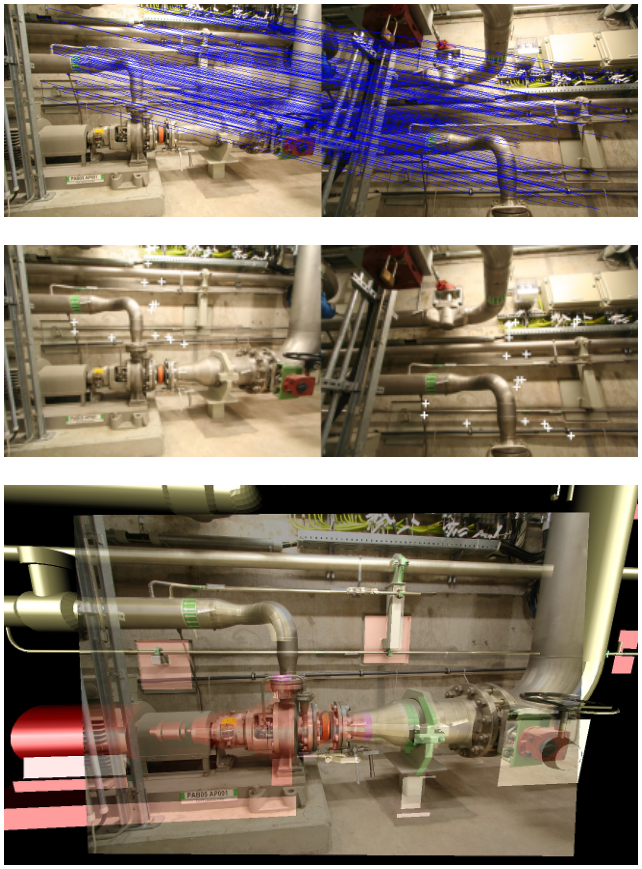

Figure 2: Scale recovery using a plane: The top graphic shows the result of the robust matching (left image is the target; the right the keyframe). The middle graphic shows: white crosses represents the points extracted from the plane. The resulting augmentation is displayed at the bottom.

References

[1] R. Cipolla and T. Drummond. Real-time tracking of complex structures with on-line camera calibration. In *BMVC*, pages 574–583, Sep. 1999.

[2] M. Fiala. ARTag, a Fiducial Marker System Using Digital Techniques. *CVPR*, 2:590–596, Jun. 2005.

[3] P. Georgel, P. Schroeder, S. Benhimane, S. Hinterstoisser, M. Appel, and N. Navab. An Industrial Augmented Reality Solution For Discrepancy Check . In *ISMAR*, pages 311–314, Nov. 2007.

[4] R. Hartley. Estimation of relative camera positions for uncalibrated cameras. In *ECCV*, pages 579–587, May 1992.

[5] V. Lepetit, L. Vacchetti, D. Thalmann, and P. Fua. Fully Automated and Stable Registration for Augmented Reality Applications. page 93, Oct. 2003.

[6] N. Navab, B. Bascle, M. Appel, and E. Cubillo. Scene augmentation via the fusion of industrial drawings and uncalibrated images with a view to marker-less calibration. In *IWAR*, pages 125–33, Oct. 1999.

[7] N. Navab, E. Cubillo, B. Bascle, J. Lockau, K. Kamsties, and M. Neuberger. CYLICON: A Software Platform for the creation and update of virtual factories. In *ETFA*, Oct. 1999.

[8] K. Pentenrieder, C. Bade, F. Doil, and P. Meier. Augmented reality-based factory planning - an application tailored to industrial needs. In *ISMAR*, pages 31–42, Nov. 2007.

[9] J. Platonov, H. Heibel, P. Meier, and B. Grollmann. A mobil markerless ar system for maintenance and repair. In *ISMAR*, pages 105–108, Oct. 2006.

[10] L. Vacchetti and V. Lepetit. Stable real-time 3d tracking using online and offline information. *PAMI*, 26(10):1385–1391, 2004.

[11] S. Webel, M. Becker, D. Stricker, and H. Wuest. Identifying differences between CAD and physical mock-ups using AR. In *ISMAR*, pages 281–282, Nov. 2007.