Combined Registration Methods for Pose Estimation

Dong Han[1]*, Bodo Rosenhahn[2], Joachim Weickert[3], and Hans-Peter Seidel[4]

[1] University of Bonn, Germany (han@ins.uni-bonn.de)
[2] University of Hannover, Germany (rosenhahn@tnt.uni-hannover.de)
[3] Saarland University, Germany (weickert@mia.uni-saarland.de)
[4] Max-Planck Institute for Informatics, Germany (hpseidel@mpi-sb.mpg.de)

Abstract. In this work, we analyze three different registration algorithms: Chamfer distance matching, the well-known iterated closest points (ICP) and an optic flow based registration. Their pairwise combination is investigated in the context of silhouette based pose estimation. It turns out that Chamfer matching and ICP used in combination do not only perform fairly well with small offset, but also deal with large offset significantly better than the other combinations. We show that by applying different optimized search strategies, the computational cost can be reduced by a factor eight. We further demonstrate the robustness of our method against simultaneous translation and rotation.

1 Introduction

Shape registration is an important technique in computer vision. It is present in many applications, such as image segmentation, object recognition and classification, motion tracking or image retrieval. The task of shape registration is to establish point-to-point correspondences between two images [1]. In the context of this paper, by registration or matching we mean to estimate the geometric transformation between the reference image and the 3D target object.

As a classic method for solving correspondence problems in computer vision, shape matching has been intensively studied in recent years, see e.g. [2, 3]. A survey is available in [4]. A very popular shape matching algorithm is the iterated closest points (ICP) algorithm [5], which uses explicit representations like points and curves. Some variants of the ICP algorithm were evaluated in [6] by comparing each stage of the algorithms and their speed of convergence. Another popular approach is to use optic flow which has been an intensive research field for decades, because of its capability for image sequence analysis. Horn and Schunck presented in [7] a global method to build dense flow field using variational framework. A performance evaluation of many well-known optic flow methods is available in [8]. The Chamfer distance matching algorithm was proposed by Barrow et al. [9]. It has the nice property of being able to deal with large

* The work was done when the author was writing his thesis at Saarland University and Max-Planck Institute for Informatics.

offsets, efficient and easy to implement. Borgefors [10, 11] improved the Chamfer matching by using a more reasonable confidence measure and embedding the basic algorithm into a resolution pyramid, which has reduced the computational cost significantly. A recent application was presented in [12], where Gavrila used a multi-feature hierarchical algorithm to match N templates simultaneously and demonstrated the application in traffic sign detection. We will explain these registration algorithms with more details in Section 2.

In most work, a performance analysis with respect to complexity or stability, especially for diverse registration methods as compared here (ICP, optic flow, Chamfer) is still missing. Therefore, the main contribution of this work is to individually analyse these approaches and to evaluate their performance in all possible combinations for different rigid motions. We further investigate different variants for Chamfer matching to improve the speed without degradation in performance. As test scenario we concentrate on 2D-3D pose estimation. By *Pose* we refer to the definition in [13] as "the transformation needed to map an object model from its own inherent coordinate system into agreement with the sensory data". In general, the task of pose estimation is to find this transformation. In the scope of this paper, we restrict the transformation to a rigid body motion. Much work on pose estimation has been done in different abstraction level of geometric descriptors (see [14] for detailed overviews).

In this work, we extend the joint pose estimation algorithm in [15] by embedding the Chamfer distance matching, to be able to deal with large movements. Section 2 summarizes the three used registration methods. The experimental setup is discussed in Section 3. Performance of the presented approaches is evaluated in Section 4. This paper is concluded with a brief summary in Section 5.

2 Registration Algorithms

2.1 Registration with ICP

The iterated closest points algorithm was introduced by Besl and McKay [5]. It is a method for aligning 3D models based on geometry, widely used for registering outputs of 3D scanners. ICP starts with two point clouds and an initial guess for their relative rigid body transformation. The basic idea is to refine this transformation and to minimize the error by iterating the following steps:

1. Give a good assumption of the initial relative pose of the object model to the reference image.
2. Find the closest points, which results in pairs of corresponding points.
3. Calculate the transformation such that an error metric is minimized [6].
4. Go to Step 2.

In [16], Zhang proposed an improved ICP algorithm to determine the correspondence set between the image contour and the 3D object pose, where he used a K-dimensional tree to partition the point sets, which reduced the computational cost of registering large image data significantly. In [15], this ICP variant was applied for a pose tracking system, in order to find the point correspondences between the image silhouette and the 3D contour.

2.2 Registration with Level Set Functions and Optic Flow

The level set method was originally introduced by Dervieux and Thomasset [17] in 1979 and became popular by a paper of Osher and Sethian [18] in 1988. Since the level set method could easily integrate further constraints like 2D or 3D shape prior, it has been very popular in image segmentation in recent years. In [19], Brox et al. introduced a segmentation algorithm integrating color, texture and motion information, where the nonlinear diffusion was used for feature extraction and the level set approach was applied for the total energy minimization.

The original idea of level set formulation is to define a smooth function $\Phi : \Omega \mapsto \mathbb{R}$, which represents a contour Γ in $\mathbb{R}^n$ as the set where $\Phi(x) = 0$, bounding an open region Ω. Then the level set function Φ for defining the image region has the following properties: $\Phi(x) > 0$ if $x \in \Omega_1$, $\Phi(x) < 0$ if $x \in \Omega_2$, and $\Phi(x) = 0$ if $x \in \Gamma$, where Ω_1 represents the object and Ω_2 represents the background. The zero-level line is the searched boundary between the two regions. This formulation has several advantages. It is invariant under topological changes of the regions, which makes it very convenient to handle occlusion. It can also be easily extended, e.g. by having further constraints like shape prior [20].

For the task of matching, the object contour C is formulated with the level set function introduced in [21]:

$$\Phi(x) = \begin{cases} D(x, C) & \text{for } x \text{ inside } C \\ -D(x, C) & \text{for } x \text{ outside } C \\ 0 & \text{for } x \in C \end{cases}$$

where $D(x, C)$ denotes the Euclidean distance of $x \in \Omega$ to the closest point $\tilde{x}$ on the contour C. Obviously this level set formulation is invariant under rigid body motion.

For matching the distance transformed images, instead of the straightforward distance measure $d(\Phi_1, \Phi_2) = \int_\Omega (\Phi_1(x) - \Phi_2(x))^2 \mathbf{dx}$, it is suggested in [21] to find the optimized transformation which minimizes the energy functional:

$$E(\tau) = \int_\Omega (\Phi_1(x) - \Phi_2(\tau x))^2 + \alpha(|\nabla u|^2 + |\nabla v|^2)\mathbf{dx} \tag{1}$$

where $\tau x = x + \mathbf{w}(x)$ and $\mathbf{w}(x) := (u(x), v(x))^\top$ is the displacement vector. α is a regularization parameter for penalizing the smoothness. Minimization of the functional yields the estimation of the shape deformation field $\mathbf{w}$, which comes down to an optic flow estimation problem. In the context of pose estimation, optic flow was computed between the distance transform of the image contour and the distance transform of the projected model contour, so as to get additional correspondences of 2D points in successive images. This improves the 3D registration.

2.3 Registration with Chamfer Distance Matching

Barrow et al. [9] proposed the Chamfer matching algorithm in 1977. The algorithm compares two images: the input image and the so-called template image.

The goal is to find the best fit of edge points between them. It has several advantages, such as the ability to handle noisy or distorted images. With a linear complexity to the number of the corresponding points, the algorithm is very fast.

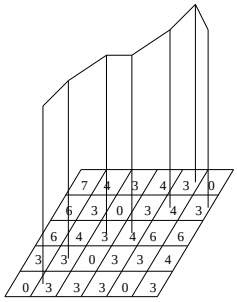

Fig. 1. The transformed template is superimposed on the distance transformed image. The distance measure is evaluated by the r.m.s. average of the pixel values that are hit. In this case, it is: $\frac{1}{3}\sqrt{\frac{1}{7}(3^2+3^2+3^2+4^2+4^2+3^2+3^2)} = 3.67$.

Assume an input image and a template image are given. Both of them are binary and pre-segmented. The images contain feature points and non-feature points. Here one does not care what exactly the specific feature is, one just separates the features from non-features. Some distance transformation (DT) [10] is applied on these images, usually the Euclidean distance transform or its approximation [11]. The transformed template is placed over the transformed input image, as is illustrated in Figure 1. This superimposition gives the possibility to measure the correspondences between the contours. Values of the pixels on the transformed input image, which the transformed template (also regarded as polygon in [11]) hits, measure exactly how the image differs from the template. To evaluate the matching, one just needs to consider this array, whose elements represent distance to the nearest feature point. The distance measurement has been improved over the last several decades. In the original Chamfer matching, the arithmetic average was chosen for the matching measure (so-called *Chamfer distance*): $D = \frac{1}{n}\sum_{i=1}^{n} v_i$, where n is the number of points in the polygon. In [11], Borgefors has chosen the root mean square average (r.m.s. for short): $D = \frac{1}{d}\sqrt{\frac{1}{n}\sum_{i=1}^{n} v_i^2}$, where d is the unit distance in the distance transform. (Note: in Figure 1, $d = 3$.)

In [10], a comparison of four different average measurements was presented: median, arithmetic average, r.m.s., and maximum. After careful investigation, it was observed that r.m.s. generated the fewest local minima that disturbed the algorithm, therefore resulting in a more accurate convergence.

3 Application: Silhouette Based 2D-3D Pose Estimation

The three used registration algorithms are tested in the context of silhouette based pose estimation. In the joint system, 3D objects (a non-convex teapot model is used in the experiments) are projected onto the image plane by a multi-view setup. The projection matrices are predefined. The images which are captured by a calibrated camera system, are processed by using an image segmentation algorithm. This procedure gives the silhouette of the considered object. The correspondence estimation between the initial pose and the segmented object silhouette is achieved by using the registration algorithms: CM [9, 11], ICP [16], OF [22] and their combinations. The resulting point correspondences are used to generate the new pose and the updated pose is once again projected. Then the new generated contour proceeds as shape prior to the segmentation procedure. The iteration between pose estimation and contour segmentation is repeated several times. Finally the estimated pose is used for the next frame. The whole process is illustrated in Figure 2.

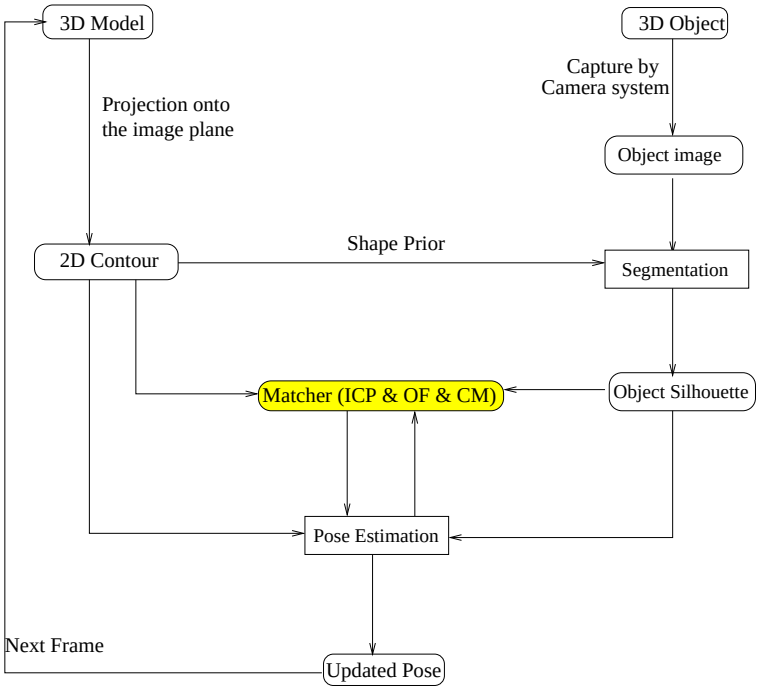

Fig. 2. Flow Chart of the silhouette based pose estimation algorithm.

In the following section, the performance of ICP, optic flow as well as the Chamfer matching are evaluated. By simultaneous usage of correspondences from different registration algorithms, their combinations are also analyzed.

4 Experiments and Evaluation

4.1 Convergence analysis of plain 3D translations

In this section, three different registration algorithms (ICP, Optic flow, and Chamfer) are compared in terms of silhouette based pose estimation. A teapot model is used for the experiments. We use a similar experimental setup as described in [21]. The model silhouette is moved in 3D to be far from the object contour in the images. To this end, we compute the contour and the pose in the first frame as usual, then we disturb the initial pose by translation and rotation.

In the first experiment, we analyze the accuracy and stability of the three registration algorithms with respect to the same initial movements in translation. Figure 3 shows certain frames taken from a stereo image sequences of the teapot. The teapot is captured from the left and right cameras. Initially the pose of the teapot is disturbed in the area of [-15,15] cm along the x, y, and z axes. Random samples of translational disturbances are generated in this interval. We test the performance of registration methods for correspondence estimation. When the difference between the estimated pose and the correct pose is within some predefined threshold, it is regarded as converged.

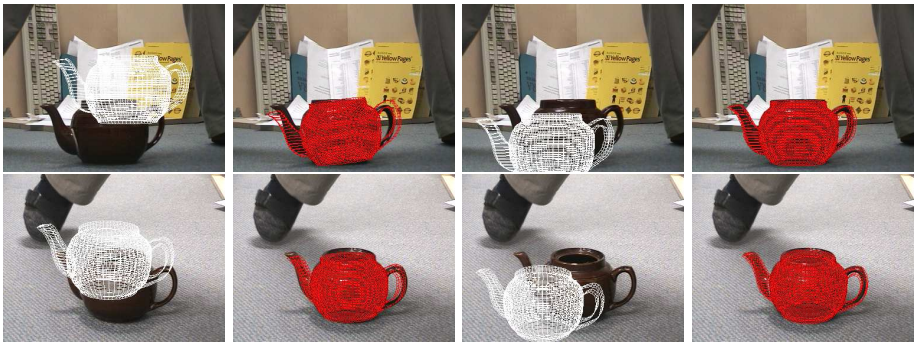

Fig. 3. Some frames from a stereo image sequences (400 frames). The first column is the initial translated image. The estimated pose is shown in the second column, which has not converged. The third column is the initialization of another translated image, where the corresponding converged result is shown in the last column. **Top row:** left view. **Bottom row:** right view.

In the left column of Figure 4, we show the correspondence clouds of 400 random samples of translational disturbances in [-15, 15] cm by invoking ICP, optic flow and Chamfer matching respectively. Blue stars denote the cases where the registration method succeeds in converging back to the correct pose, while red crosses denote the unconverged cases. Regarding that ICP becomes unstable when the initial pose is disturbed too much (as shown in the top left image), Chamfer matching is much better at dealing with large movements in translation.

In the bottom left image, it is shown that, with an initial movement of as large as [-15, 15] cm, a very wide area has converged.

For this reason, we expand the correspondences near the spout of the teapot in a nearly converged frame, in order to tell the different properties between these matchers. They are visualized in Figure 6. The small line segments between the contour points and estimated corresponding points are sketched. They are generated by optic flow, ICP, and Chamfer matching respectively. One can see that, in the middle image the line segment is perpendicular to the object silhouette, i.e. the object is still moving towards its exact location and has not yet converged. Therefore we conclude that ICP is good at dealing with tiny offsets and improving the pose result subtly. While the object in the right image has already converged, so that Chamfer matching at this stage would not give further improvement any more. Therefore we are motivated to augment the capability of Chamfer matching with ICP and to demonstrate they are actually complementary to each other.

In the second experiment, we evaluate the performance of different combination of registration algorithms at dealing with large movements in translation. These algorithms can easily be combined in the sense of adding up their respective estimated point correspondences. The right column of Figure 4 shows the correspondence clouds applying certain combination of ICP, optic flow and Chamfer matching. Table 1 shows the convergence rates of all the three different matchers and their combinations. The combination of ICP and CM gives the best result. 95.3% of the random instances have converged with an initial movement of [-15, 15] cm. After further experiments, we observed that, even with movement of [-25, 25] cm, about 60% convergence rate could be achieved by applying CM and ICP simultaneously. Table 2 shows the decreasing convergence rate with the increasing movement from [-15, 15] cm till [-25, 25] cm.

Table 1. Comparison of convergence rates (of the correspondence clouds from Fig. 4).

Matcher	Convergence Rate (translation)
ICP	30.6%
OF	46.2%
CM	51.3%
ICP + OF	47.3%
CM + ICP	**95.3**%
CM + OF	81.8%
ICP + OF + CM	81.5%

4.2 Variants for Chamfer matching

In the Chamfer distance matching algorithm, we define an operation called *Co-convolve*, which takes two matrices and goes through the larger-size matrix

by superimposing the smaller-size matrix at all possible locations. The boundary area is filled with zero if necessary. The whole image has been involved in the Co-convolution process. Large images slow down the algorithm. In our setup (2×2.33 GHz AMD Opteron processor, C++ code, image size: 376×284) we achieved 11 frames per hour. The slow convergence is partly due to the unessential regions within the image, which also deteriorates the convergence rate. One way to speed up the method is to omit the construction of the boundary pixels of the template image, namely just to convolve its central part. Then the observed speed is four times as fast as before, but meanwhile the convergence rate decreases enormously. Therefore a new stopping criterion is implemented. A local search (as illustrated in Figure 5) is performed during Co-convolution and is forced to stop when the best matching location is not improved during five consecutive circulations. This has shown to increase the speed dramatically. The algorithm has a speed of 70 frames per hour while invoking only Chamfer matching in the image registration part. The combination of Chamfer matching and ICP gives again the best result.

Table 2. Convergence rates of applying CM and ICP simultaneously with different initial movements.

Matcher	±15cm	±17.5cm	±20cm	±22.5cm	±25cm
CM + ICP	95.3%	84%	78%	69.6%	63.2%

Table 3. Comparison of convergence rates of the original Chamfer matching algorithm; its simplified version, which only convolves the central part; an improved version with more accurate stopping criterion; and the version with another search strategy. All the percent numbers stand for the convergence rates.

	±7.5cm	±15cm	±10cm (& ICP)	±15cm (& ICP)	speed
CM(original)	88.8%	47.9%	> 95%	94.8%	11 f/h
CM(central part)	56.8%	27.8%	56.7%	39.9%	40 f/h
CM(new stopping)	87.8%	51.3%	98.8%	95.3%	70 f/h
CM(new search)	79.3%	48.5%	94.5%	77.8%	80 f/h

As an alternative, another search strategy for Chamfer matching has been implemented. The method starts from the initial location as usual, which is considered as seed. Then the eight neighbors of the seed are evaluated. The best matching location among the seed and its eight neighbors is regarded as the new seed. If this best matching location is the seed itself, the search is forced to stop. By this means, the search is always progressing towards the best matching direction and makes it even faster. It has a speed of about 80 frames per hour. Table 3 compares the speed, convergence rate of applying CM alone and that of applying CM and ICP together among these four different cases.

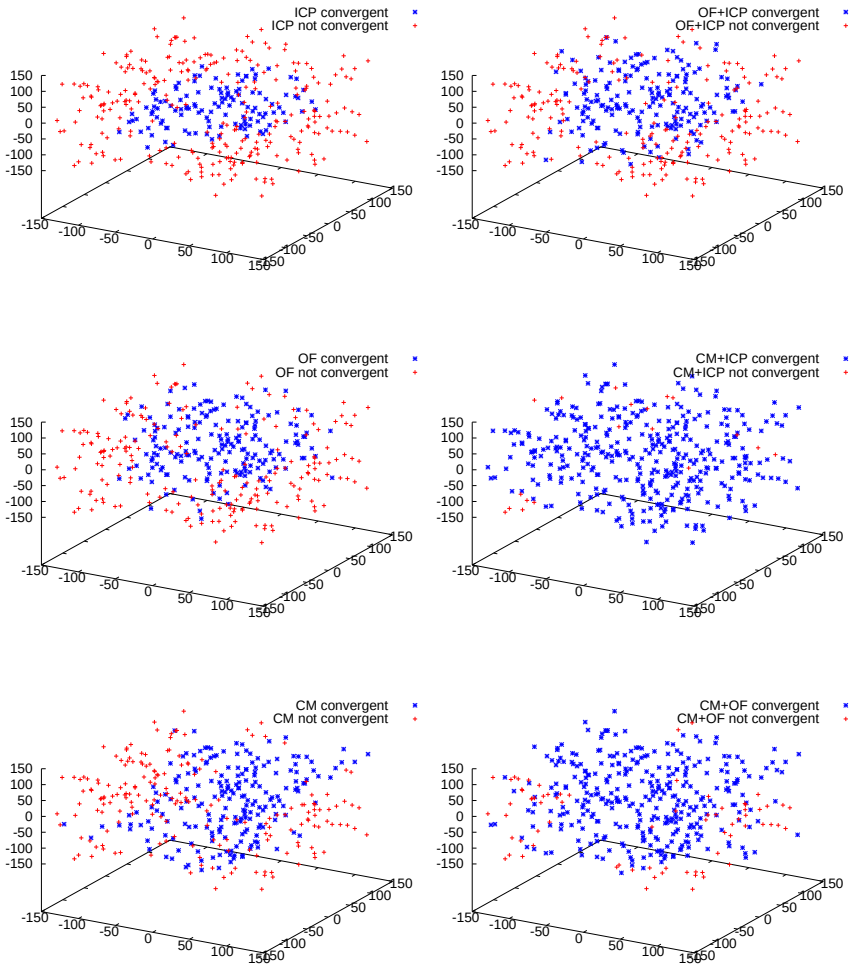

Fig. 4. With random disturbance in the area of [-15, 15] cm in all three dimensions. The left column is the convergence performance of solely using ICP, Optic Flow and CM respectively. The right column is the convergence performance of invoking combination of ICP and OF, combination of CM and ICP, and combination of CM and OF respectively. See text for more details. This figure is best viewed in color.

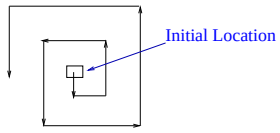

Fig. 5. Illustration of the local search during Co-convolution.

Fig. 6. The spout of the teapot from the right view of a nearly converged frame. **From left to right**: OF, ICP and Chamfer Matching respectively.

4.3 Combined rotations and translations

Whereas in the above experiments only translational movements are analyzed, we now switch to simultaneous rotations and translations. In addition to the initial movements in translation, movements in rotation around the x, y, and z axes are performed. For each image registration algorithm and their combinations, we evaluate the performance dealing with simultaneous translation and rotation. The convergence rate of each instance of the matching algorithms is shown in a separate diagram in Figure 7 respectively. For each instance, 20 distinct initializations (all possible transformations by combining movements in translation $\{\pm 7.5, \pm 10, \pm 12.5, \pm 15, \pm 17.5\,(\text{cm})\}$ and movements in rotation $\{\pm 5, \pm 10, \pm 15, \pm 20\,(\text{degrees})\}$) are investigated in the experiments, where the individual results are connected with lines. x-coordinates are scaled by 2, since we actually have movements in the whole interval, e.g. $\pm$15cm represents [-15, 15] cm. From the surfaces constructed by these segments, we can see in a concrete situation, which combination of the matchers performs best.

Generally, the single ICP algorithm performs quite well at dealing with both translations and rotations. However, the accuracy is deteriorating with increasing movements. Combined with optic flow algorithm, the precision of ICP decreases less rapidly. Obviously the combination of ICP and Chamfer matching improves the stability and convergence behavior of the pose estimation algorithm significantly. A convergence rate of over 60% can be achieved even when the pose of the object model is disturbed in the area of [-17.5, 17.5] cm along the axes and [-20, 20] degrees around the axes.

5 Conclusions

In this work we individually analyze three registration methods (Chamfer, ICP and optic flow) and respective combinations in their performance for 2D-3D pose estimation. The experiments reveal, that the methods have different advantages and shortcomings, but can efficiently be combined to gain a registration algorithm which can handle large displacements in reasonable time and accuracy. After optimizing, our proposed joint algorithm works in 70 to 80 frames per hour on a standard PC, which increased the efficiency by factor 7, compared to

the original implementation. The evaluation results also reveal, that it is not always advantageous to combine in a naive way all available registration methods (here ICP, Chamfer and OF). Instead, reasonable combination of registration methods which compensate for each other's drawbacks (here Chamfer and ICP) can be much better for both performance and convergence.

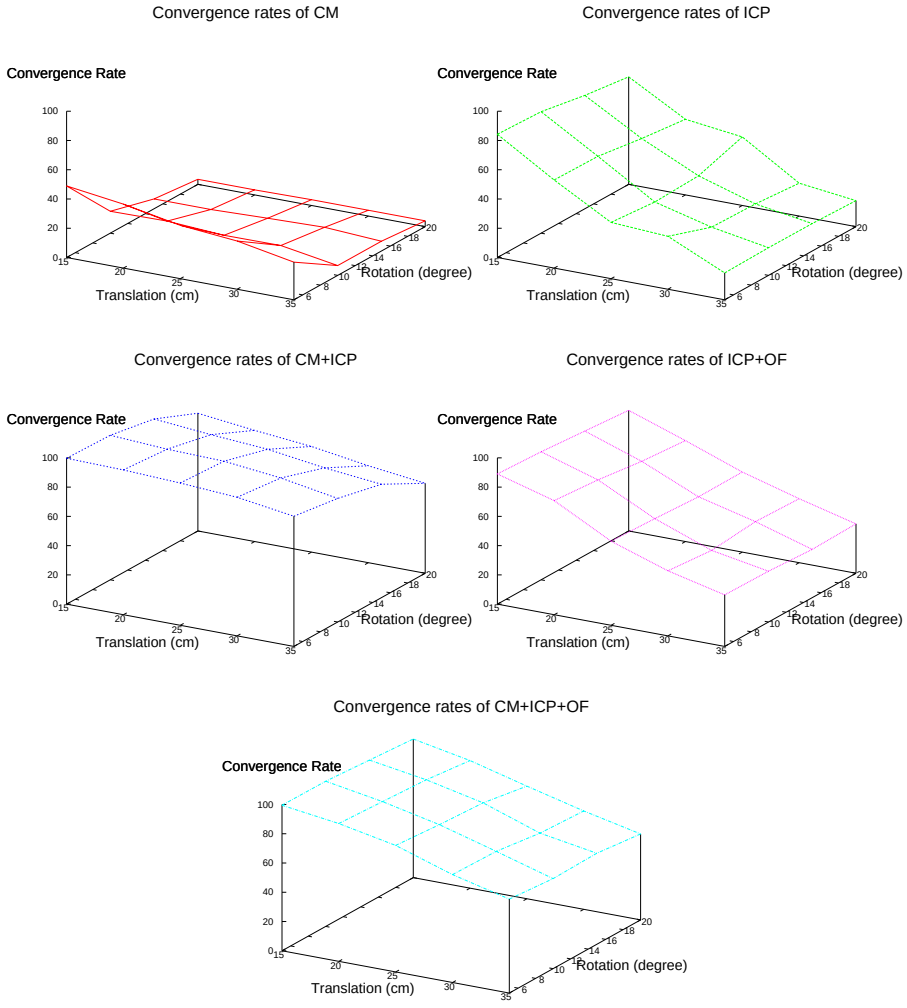

Fig. 7. From left to right, from top to bottom: convergence rates of CM; ICP; CM+ICP; ICP+OF; and CM+ICP+OF respectively. The x-,y- and z-coordinates are initial movement in translation, initial movement in rotation and convergence rate in percent respectively. This figure is best viewed in color.

References

1. Goshtasby, A.: Tutorial on 2-d and 3-d image registration (2004) Accessed:1st July `http://www.cs.wright.edu/~agoshtas/CVPR04_Registration_Tutorial.html`.
2. Latecki, L.J., Lakamper, R.: Shape similarity measure based on correspondence of visual parts. IEEE Trans. Pattern Anal. Mach. Intell. **22** (2000) 1185–1190
3. Paragios, N., Rousson, M., Ramesh, V.: Non-rigid registration using distance functions. Comput. Vis. Image Underst. **89** (2003) 142–165
4. Veltkamp, R., Hagedoorn, M.: State-of-the-art in shape matching. Technical Report UU-CS-1999-27, Utrecht University, the Netherlands (1999)
5. Besl, P., McKay, N.: A method for registration of 3d shapes. IEEE Trans. Pattern Anal. Mach. Intell. **14** (1992) 239–256
6. Rusinkiewicz, S., Levoy, M.: Efficient variants of the ICP algorithm. In: Proc. 3rd Int'l Conf. on 3D Digital Imaging and Modeling. (2001) 145–152
7. Horn, B., Schunck, B.: Determining optical flow. Art. Intel. **16** (1981) 185–203
8. Barron, J., Fleet, D., Beauchemin, S.: Performance of optical flow techniques. International Journal of Computer Vision **12** (1994) 43–77
9. Barrow, H., Tenenbaum, J., Bolles, R., Wolf, H.: Parametric correspondence and chamfer matching: Two new techniques for image matching. In: International joint Conference on Artificial Intelligence. (1977) 659–663
10. Borgefors, G.: An improved version of the chamfer matching algorithm. In: 7th Intl Conf. on Pattern Recognition, Montreal, Canada (1984) 1175–1177
11. Borgefors, G.: Hierarchical chamfer matching: A parametric edge matching algorithm. IEEE Trans. Pattern Anal. Mach. Intell. **10** (1988) 849–865
12. Gavrila, D.: Multi-feature hierarchical template matching using distance transforms. In: IEEE Int'l Conf. on Pattern Recognition, Brisbane, Australia (1998)
13. Grimson, W.: Object recognition by computer: the role of geometric constraints. MIT Press, Cambridge, MA, USA (1990)
14. Goddard, J.: Pose and Motion Estimation From Vision Using Dual Quaternion Based Extended Kalman Filtering. PhD thesis, University of Tennessee (1997)
15. Rosenhahn, B., Brox, T., Weickert, J.: Three-dimensional shape knowledge for joint image segmentation and pose tracking. IJCV **73** (2007) 243–262
16. Zhang, Z.: Iterative point matching for registration of free-form curves and surfaces. Int. J. Comput. Vision **13** (1994) 119–152
17. Dervieux, A., Thomasset, F.: A finite element method for the simulation of a rayleigh-taylor instability. In Rautman, R., ed.: Approximation methods for Navier-Stokes problems. Volume 771., Springer (1979) 145–158
18. Osher, S., Sethian, J.: Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations. J. Comput. Phys. **79** (1988) 12–49
19. Brox, T., Rousson, M., Deriche, R., Weickert, J.: Unsupervised segmentation incorporating colour, texture, and motion. In Petkov, N., Westenberg, M.A., eds.: Computer Analysis of Images and Patterns, Springer (2003) 353–360
20. Cremers, D.: Dynamical statistical shape priors for level set based tracking. IEEE Trans. Pattern Anal. Mach. Intell. **28** (2006) 1262–1273
21. Rosenhahn, B., Brox, T., Cremers, D., Seidel, H.P.: A comparison of shape matching methods for contour based pose estimation. In: Proc. International Workshop on Combinatorial Image Analysis, Berlin, Springer (2006) 263–276
22. Brox, T., Bruhn, A., Papenberg, N., Weickert, J.: High accuracy optical flow estimation based on a theory for warping. In: European Conference on Computer Vision, Prague, Czech Republic (2004) 25–36