33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

ECM versus ICP for Point Registration

Weiguo Xie, Lutz-Peter Nolte, and Guoyan Zheng

Abstract—**Iterative Closest Point (ICP) is a widely exploited method for point registration that is based on binary point-to-point assignments, whereas the Expectation Conditional Maximization (ECM) algorithm tries to solve the problem of point registration within the framework of maximum likelihood with point-to-cluster matching.**

In this paper, by fulfilling the implementation of both algorithms as well as conducting experiments in a scenario where dozens of model points must be registered with thousands of observation points on a pelvis model, we investigated and compared the performance (e.g. accuracy and robustness) of both ICP and ECM for point registration in cases without noise and with Gaussian white noise.

The experiment results reveal that the ECM method is much less sensitive to initialization and is able to achieve more consistent estimations of the transformation parameters than the ICP algorithm, since the latter easily sinks into local minima and leads to quite different registration results with respect to different initializations. Both algorithms can reach the high registration accuracy at the same level, however, the ICP method usually requires an appropriate initialization to converge globally. In the presence of Gaussian white noise, it is observed in experiments that ECM is less efficient but more robust than ICP.

I. Introduction

THE Iterative Closest Point (ICP) method as well as its numerous variants are widely used for point registration [1]-[2]. It alternates between binary point-to-point correspondence searching and optimal transformation estimation, and works efficiently [3]. However, due to its strict binary selection of point-to-point assignments, ICP is easily trapped in local minima and thus, is quite sensitive to initializations and the acceptance/rejection threshold of an assignment [3]. For instance, the ICP algorithm is employed in [2] for point registration, where the inconsistency of registration results with respect to different initializations is noticeable, because of the existence of different local minima.

Probabilistic methods are other important options for point registration. By utilizing generally a Gaussian Mixture Model (GMM), they transform the point-to-point matching problem into the scope of maximum likelihood with missing data (the point-to-cluster assignments) [3]. Although the Expectation Maximization (EM) algorithm can be chosen to solve such a maximum likelihood problem, i.e. estimation of the mixture parameters and the missing data, casting of point registration problems in EM framework remains intrinsically difficult [3].

Therefore, Horaud et al. proposed an Expectation Conditional Maximization (ECM) algorithm instead of the EM method to solve the point registration problem within the framework of maximum likelihood with missing data. In ECM, each M-step is replaced by a sequence of conditional maximization steps (CM-steps), and these conditional maximizations over the registration parameters cannot be carried out independently of other parameters, namely covariances of the GMM model [3]. It ensures that the algorithm does not become easily stuck in local minima as ICP does.

In this paper, we implement both ICP and ECM algorithms and compare their performance in a scenario in which dozens of model points have to be matched with thousands of observation points. We will evaluate their registration accuracies defined by the average of distances between model points transformed by the estimated transformation parameters and those transformed by the ground truth of the transformation. By adding Gaussian white noise with the mean centered at each observation point and a certain standard deviation, we observe also the robustness of both algorithms.

II. Methods

A. ECM for Point Registration

The ECM algorithm we implemented and compared with the ICP algorithm originates from [3].

Assume we have 3D observation points $\{\boldsymbol{Y}_j\}_{1\le j\le m}$ and 3D model points $\{\boldsymbol{X}_i\}_{1\le i\le n}$, then the transformed model point can be denoted as $\boldsymbol{\mu}(\boldsymbol{X}_i;\boldsymbol{\theta}) = \boldsymbol{R}\boldsymbol{X}_i + \vec{\boldsymbol{t}}$ with transformation parameters $\boldsymbol{\theta} := \{\boldsymbol{R}, \vec{\boldsymbol{t}}\}$ where $\boldsymbol{R}$ is a 3×3 rotation matrix and $\vec{\boldsymbol{t}}$ is a 3×1 translation vector [3].

The prior probability that an observed point $\boldsymbol{Y}_j$ is assigned to a Gaussian cluster with center $\boldsymbol{\mu}(\boldsymbol{X}_i;\boldsymbol{\theta})$ can be expressed by $p_i = P(Z_j = i)$ and the prior probability that $\boldsymbol{Y}_j$ corresponds to an outlier is written as $p_{n+1} = P(Z_j = n+1)$ [3].

Moreover, the conditional likelihood of $\boldsymbol{Y}_j$, i.e. the probability of $\boldsymbol{Y}_j$ given its cluster assignment, is $P(\boldsymbol{Y}_j|Z_j = i)\ \forall j \in \{1,\dots,m\}\ and\ \forall i \in \{1,\dots,n+1\}$ that follows a Gaussian distribution with mean $\boldsymbol{\mu}(\boldsymbol{X}_i;\boldsymbol{\theta})$ and covariance $\boldsymbol{\Sigma}_i$.

Manuscript received April 15, 2011.

Weiguo Xie is with the Institute for Surgical Technology and Biomechanics (ISTB), University of Bern, Stauffacherstr. 78, CH-3014 Bern, Switzerland (phone: +41(0)31-631-5928; e-mail: weiguo.xie@istb.unibe.ch).

Lutz-Peter Nolte is with the Institute for Surgical Technology and Biomechanics (ISTB), University of Bern, Stauffacherstr. 78, CH-3014 Bern, Switzerland (e-mail: Lutz.Nolte@MEMcenter.unibe.ch).

Guoyan Zheng is with the Institute for Surgical Technology and Biomechanics (ISTB), University of Bern, Stauffacherstr. 78, CH-3014 Bern, Switzerland (e-mail: guoyan.zheng@istb.unibe.ch).

Similarly, the conditional likelihood of $\boldsymbol{Y}_j$ belonging to the outlier cluster is a uniform distribution over a 3D volume V: $P(\boldsymbol{Y}_j|Z_j = n+1) = \frac{1}{V}$.

Thus the posterior probabilities of an assignment $P(Z_j = i)$ conditioned by observations can be derived as:

$$\alpha_{ji} = P(Z_j = i|\boldsymbol{Y}_j) = \frac{P(\boldsymbol{Y}_j|Z_j=i)P(Z_j=i)}{P(\boldsymbol{Y}_j)} = \frac{P(\boldsymbol{Y}_j|Z_j=i)P(Z_j=i)}{\sum_i^{n+1} P(\boldsymbol{Y}_j|Z_j=i)P(Z_j=i)} = \frac{P(\boldsymbol{Y}_j|Z_j=i)p_i}{\sum_i^{n+1} P(\boldsymbol{Y}_j|Z_j=i)p_i} = \frac{|\boldsymbol{\Sigma}_i|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\|\boldsymbol{Y}_j - \boldsymbol{\mu}(\boldsymbol{X}_i;\boldsymbol{\theta})\|^2_{\boldsymbol{\Sigma}_i}\right)}{\sum_{i=1}^{n} |\boldsymbol{\Sigma}_i|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\|\boldsymbol{Y}_j - \boldsymbol{\mu}(\boldsymbol{X}_i;\boldsymbol{\theta})\|^2_{\boldsymbol{\Sigma}_i}\right) + 1.5\sqrt{2\pi}r^{-3}}, \quad (1)$$

where r is radius of a small sphere centered at a model point $\boldsymbol{X}_i$ and $v = \frac{4}{3}\pi r^3$ is the sphere volume with assumption of $nv \ll V$ and $p_i = \frac{v}{V}, 1 \le i \le n$ or $p_i = \frac{V-nv}{V}, i = n+1$, which is equation (12) of [3].

The expected complete-data log-likelihood conditioned by the observed data is given by equation (18) in [3]:

$$\boldsymbol{\varepsilon}(\boldsymbol{\psi}) = -\frac{1}{2}\sum_{j=1}^{m}\sum_{i=1}^{n} \alpha_{ji}\left(\|\boldsymbol{Y}_j - \boldsymbol{\mu}(\boldsymbol{X}_i;\boldsymbol{\theta})\|^2_{\boldsymbol{\Sigma}_i} + \log|\boldsymbol{\Sigma}_i|\right) \quad (2)$$

with $\boldsymbol{\psi} = \{\boldsymbol{\theta}, \boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_n\}$.

The estimation of transformation parameter $\boldsymbol{\theta}$ from (2) is separated from and ahead of the estimation of covariance matrices $\boldsymbol{\Sigma}_i$ of GMM, i.e. equation (19) of [3]

$$\Theta^* = \text{argmin}_{\boldsymbol{\theta}} \frac{1}{2}\sum_{j=1}^{m}\sum_{i=1}^{n} \alpha_{ji}\|\boldsymbol{Y}_j - \boldsymbol{\mu}(\boldsymbol{X}_i;\boldsymbol{\theta})\|^2_{\boldsymbol{\Sigma}_i}. \quad (3)$$

With the introduced virtual observation points $\boldsymbol{W}_i (1 \le i \le n)$ from equation (22) of [3] and their weights $\lambda_i (1 \le i \le n)$ from equation (23) of [3], (3) can be rewritten as equation (26) in [3]

$$\Theta^* = \text{argmin}_{\boldsymbol{R},\vec{\boldsymbol{t}}} \frac{1}{2}\sum_{i=1}^{n} \lambda_i\|\boldsymbol{W}_i - \boldsymbol{R}\boldsymbol{X}_i - \vec{\boldsymbol{t}}\|^2_{\boldsymbol{\Sigma}_i}. \quad (4)$$

Supposing GMM is an isotropic covariance model ($\boldsymbol{\Sigma}_i = \sigma_i^2\boldsymbol{I}_3$), we can write (4) further as:

$$\Theta^* = \text{argmin}_{\boldsymbol{R},\vec{\boldsymbol{t}}} \frac{1}{2}\sum_{i=1}^{n} \frac{\lambda_i}{\sigma_i^2}\|\boldsymbol{W}_i - \boldsymbol{R}\boldsymbol{X}_i - \vec{\boldsymbol{t}}\|^2 = \text{argmin}_{\boldsymbol{R},\vec{\boldsymbol{t}}} \frac{1}{2}\sum_{i=1}^{n} \left\|\frac{\sqrt{\lambda_i}}{\sigma_i}\boldsymbol{W}_i - \frac{\sqrt{\lambda_i}}{\sigma_i}(\boldsymbol{R}\boldsymbol{X}_i + \vec{\boldsymbol{t}})\right\|^2. \quad (5)$$

Referring to [4], both centroids of weighted virtual observation points and weighted model points are respectively

$$p' \triangleq \frac{1}{n}\sum_{i=1}^{n} \frac{\sqrt{\lambda_i}}{\sigma_i}\boldsymbol{W}_i, \quad \text{and } p \triangleq \frac{1}{n}\sum_{i=1}^{n} \frac{\sqrt{\lambda_i}}{\sigma_i}\boldsymbol{X}_i.$$

The centroid of weighted transformed model points is

$$p'' \triangleq \frac{1}{n}\sum_{i=1}^{n} \frac{\sqrt{\lambda_i}}{\sigma_i}p_i'' = \frac{1}{n}\sum_{i=1}^{n} \frac{\sqrt{\lambda_i}}{\sigma_i}(\boldsymbol{R}\boldsymbol{X}_i + \vec{\boldsymbol{t}}) = \boldsymbol{R}p + \frac{1}{n}\sum_{i=1}^{n} \frac{\sqrt{\lambda_i}}{\sigma_i}\vec{\boldsymbol{t}}.$$

Denoting $\frac{1}{n}\sum_{i=1}^{n} \frac{\sqrt{\lambda_i}}{\sigma_i}$ as a, from $p' = p''$ we derive

$$p' = \boldsymbol{R}p + a\vec{\boldsymbol{t}},$$

then we obtain

$$\vec{\boldsymbol{t}} = (p' - \boldsymbol{R}p)/a. \quad (6)$$

With (6), the cost function of (5) can be written as:

$$\Theta^* = \text{argmin}_{\boldsymbol{R},\vec{\boldsymbol{t}}} \frac{1}{2}\sum_{i=1}^{n} \left\|\frac{\sqrt{\lambda_i}}{\sigma_i}\boldsymbol{W}_i - \frac{\sqrt{\lambda_i}}{\sigma_i}\boldsymbol{R}\boldsymbol{X}_i - \frac{\sqrt{\lambda_i}}{\sigma_i}\frac{(p' - \boldsymbol{R}p)}{a}\right\|^2 = \text{argmin}_{\boldsymbol{R},\vec{\boldsymbol{t}}} \frac{1}{2}\sum_{i=1}^{n} \left\|\frac{\sqrt{\lambda_i}}{\sigma_i}\left(\boldsymbol{W}_i - \frac{p'}{a}\right) - \frac{\sqrt{\lambda_i}}{\sigma_i}\boldsymbol{R}\left(\boldsymbol{X}_i - \frac{p}{a}\right)\right\|^2 = \text{argmin}_{\boldsymbol{R},\vec{\boldsymbol{t}}} \frac{1}{2}\sum_{i=1}^{n} \|q_i' - \boldsymbol{R}q_i\|^2, \quad (7)$$

where $q_i' = \frac{\sqrt{\lambda_i}}{\sigma_i}(\boldsymbol{W}_i - \frac{p'}{a})$ and $q_i = \frac{\sqrt{\lambda_i}}{\sigma_i}(\boldsymbol{X}_i - \frac{p}{a})$.

According to [4], minimizing the cost function in (7) has an equal effect as maximizing $\sum_{i=1}^{n}(q_i')^t\boldsymbol{R}q_i$. If we define $H = \sum_{i=1}^{n} q_i(q_i')^t$ whose Singular Value Decomposition (SVD) is $H = U\Lambda V^t$, we get

$$\boldsymbol{R} = VU^t. \quad (8)$$

From (8) and (6), we can estimate the new transformation parameters $\boldsymbol{R}$ and $\vec{\boldsymbol{t}}$.

When Θ^* is achieved by minimization, the partial derivative of (2) with respect to $\boldsymbol{\Sigma}_i$ is 0, i.e.

$$\frac{\partial\boldsymbol{\varepsilon}(\boldsymbol{\psi})}{\partial\boldsymbol{\Sigma}_i} = 0. \quad (9)$$

Because

$$\frac{\partial\boldsymbol{\varepsilon}(\boldsymbol{\psi})}{\partial\boldsymbol{\Sigma}_i} = -\frac{1}{2}\sum_{j=1}^{m} \alpha_{ji}\frac{\partial\left[\left(\boldsymbol{Y}_j - \boldsymbol{\mu}(\boldsymbol{X}_i;\boldsymbol{\theta})\right)^T\boldsymbol{\Sigma}_i^{-1}\left(\boldsymbol{Y}_j - \boldsymbol{\mu}(\boldsymbol{X}_i;\boldsymbol{\theta})\right) + \log|\boldsymbol{\Sigma}_i|\right]}{\partial\boldsymbol{\Sigma}_i} = -\frac{1}{2}\sum_{j=1}^{m} \alpha_{ji}\left[-\boldsymbol{\Sigma}_i^{-1}\left(\boldsymbol{Y}_j - \boldsymbol{\mu}(\boldsymbol{X}_i;\boldsymbol{\theta})\right)\left(\boldsymbol{Y}_j - \boldsymbol{\mu}(\boldsymbol{X}_i;\boldsymbol{\theta})\right)^T\boldsymbol{\Sigma}_i^{-1} + \boldsymbol{\Sigma}_i^{-1}\right], \quad (9)$$

leads to equation (20) in [3]. In case of the isotropic covariance model, (9) results in

$$\sigma_i^2 = \frac{\sum_{j=1}^{m} \alpha_{ji}\left(\boldsymbol{Y}_j - \boldsymbol{\mu}(\boldsymbol{X}_i;\boldsymbol{\theta})\right)^t\left(\boldsymbol{Y}_j - \boldsymbol{\mu}(\boldsymbol{X}_i;\boldsymbol{\theta})\right)}{3\sum_{j=1}^{m} \alpha_{ji}} = \frac{\sum_{j=1}^{m} \alpha_{ji}(\boldsymbol{Y}_j - (\boldsymbol{R}\boldsymbol{X}_i + \vec{\boldsymbol{t}}))^t(\boldsymbol{Y}_j - (\boldsymbol{R}\boldsymbol{X}_i + \vec{\boldsymbol{t}}))}{3\sum_{j=1}^{m} \alpha_{ji}}. \quad (10)$$

Thus, with new transformation parameters $\boldsymbol{R}$ and $\vec{\boldsymbol{t}}$, as well as the new posterior probabilities α_{ji}, new covariance matrices $\boldsymbol{\Sigma}_i$ can be estimated from (10).

As a result of (6), (8) and (10), the estimation of the transformation and the GMM are separately updated step by step, until the expected complete-data log-likelihood $\boldsymbol{\varepsilon}(\boldsymbol{\psi})$ conditioned by observations converges to a minimum.

B. Algorithm of ECM for Point Registration

Our implementation of the ECM algorithm consists of the following steps:

1) Initialization.
   Initialize rotation matrix with e.g. $\boldsymbol{R} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, translation vector with $\vec{\boldsymbol{t}} = \vec{\boldsymbol{0}}$, and covariance matrices with $\boldsymbol{\Sigma}_i = \sigma_i^2\boldsymbol{I}$, $i = 1, \dots, n$ and e.g. $\sigma_i^2 = 200^2(mm^2)$, assuming the isotropic covariance model is used.
2) E-step: expectation evaluation.
   (a) Use current $\boldsymbol{R}$, $\vec{\boldsymbol{t}}$ and $\boldsymbol{\Sigma}_i$ to evaluate the posterior probabilities α_{ji} from (1), virtual observation points $\boldsymbol{W}_i (1 \le i \le n)$ from equation (22) of [3] and their weights $\lambda_i (1 \le i \le n)$ from equation (23) of [3].

(b) One difference from the ECM algorithm explained in [3] is:
In order to accelerate the algorithm convergence, while the algorithm converges to a certain small extent (e.g. $\|\boldsymbol{R}^{new} - \boldsymbol{R}\| < 10^{-2}$), we utilize the "winner-takes-all" scheme to take $\boldsymbol{W}_i$ directly from the real observation point that has the biggest posterior probability with respect to a certain model point $\boldsymbol{X}_i$ than other observations, instead of to calculate $\boldsymbol{W}_i$ virtually from all weighted observation points with the equation (22) of [3].

3) CM-steps: conditional maximization.
(a) Another difference from the ECM algorithm in [3] is that we update the transformation parameters $\boldsymbol{R}$ and $\vec{\boldsymbol{t}}$ not with Semi-Definite Positive (SDP) relaxation but from (8) and (6) respectively.
(b) As mentioned above, the covariance matrices $\boldsymbol{\Sigma}_i$ is estimated separately from $\boldsymbol{R}$ and $\vec{\boldsymbol{t}}$. At this step, we can estimate them from (10) with new values of $\boldsymbol{R}$, $\vec{\boldsymbol{t}}$ and posterior possibilities α_{ji}.

4) Check convergence.
If $\|\boldsymbol{R}^{new} - \boldsymbol{R}\|$ is less than a certain convergence criterion (e.g.1e-5), go to the next step. Otherwise, the algorithm goes back to step 2.

5) Classification.
For each observation point, choose the model point with the maximal posterior probability as its matched point.

C. Standard ICP

According to [2], the standard ICP algorithm we implemented comprises procedures as follows:

1) Initialization.
Initialize the rotation matrix $\boldsymbol{R}$ and translation vector $\vec{\boldsymbol{t}}$ with the same initial values as those in the ECM algorithm we implemented above.

2) Closest point searching.
Transform the original model points $\boldsymbol{X}_i (1 \le i \le n)$ with current $\boldsymbol{R}$ and $\vec{\boldsymbol{t}}$. For each transformed model point, calculate the closest point among all observations.

3) Estimation.
Take the original model points $\boldsymbol{X}_i (1 \le i \le n)$ and their found observation correspondences as paired points to estimate new $\boldsymbol{R}$ and $\vec{\boldsymbol{t}}$ via Arun's solution in [4].

4) Check convergence.
If $\|\boldsymbol{R}^{new} - \boldsymbol{R}\|$ is less than the same convergence criterion as that of our ECM implementation, the algorithm ends. Otherwise, it goes back to step 2.

III. Experiment and Results

A. Experiment Design

The scenario we set for our experiments is that dozens of model points must be registered with significantly more observations e.g. thousands of observed points. The observation points in our experiments are from a pelvis model that is segmented from the CT dataset of a dry cadaver pelvis. As Fig. 1 shows, the 24994 3D points on this pelvis model will be considered as observed points for our experiments.

Model points used in each trial are 50 3D points that are randomly chosen from 24994 observed points and then transformed by a certain 3D transformation.

The ground truth of the 3D transformation with which the transformed model points match to their original positions on the pelvis model is known to be

$$T = \begin{bmatrix} -0.9716 & -0.1807 & 0.1520 & 158.6205 \\ 0.2077 & -0.9603 & 0.1862 & -248.4395 \\ 0.1124 & 0.2124 & 0.9707 & 14.3653 \\ 0 & 0 & 0 & 1.0000 \end{bmatrix},$$

which corresponds to Euler angles ($\alpha = 27.9°$, $\beta = 13.9°$, $\gamma = 140.8°$) and translation ($t_x = 158.62(mm)$, $t_y = -248.44(mm)$, $t_z = 14.37(mm)$).

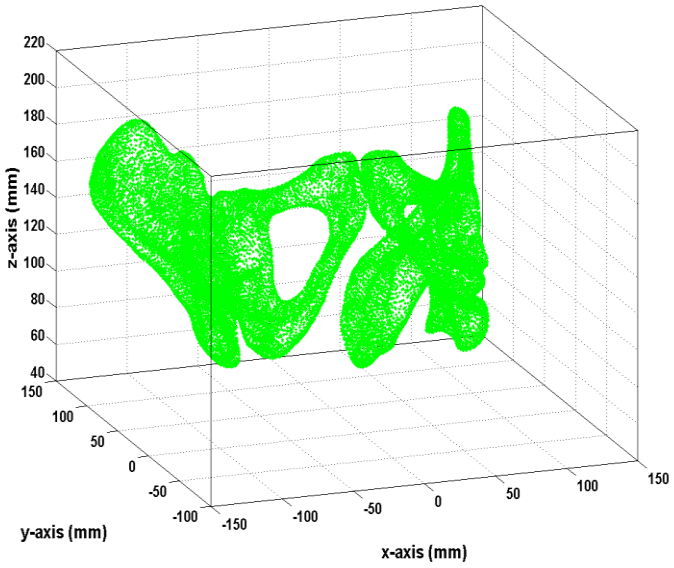

Fig. 1. The pelvis model containing 24994 3D points that serve as observed points for our experiments.

The registration accuracy in each experiment is defined by the average of distances between model points transformed by the estimated transformation parameters and those transformed by the ground truth of the transformation, i.e. $\frac{\sum_{i=1}^{n} \|(\boldsymbol{R}'\boldsymbol{X}_i + \vec{\boldsymbol{t}}') - (\boldsymbol{R}\boldsymbol{X}_i + \vec{\boldsymbol{t}})\|}{n}$, where $\boldsymbol{R}'$ and $\vec{\boldsymbol{t}}'$ are the estimated rotation and translation, $\boldsymbol{R}$ and $\vec{\boldsymbol{t}}$ are the ground truth, $n = 50$ is the number of model points, and $\boldsymbol{X}_i$ denotes a model point.

The definition of a correct registration is that its registration accuracy is better than 2mm.

Our experiments are composed of two groups: trials employing the observation points without noise, and trials with the observation points corrupted by Gaussian white noise.

The group without noise includes 40 trials for each algorithm. All of 80 trials in this group have the same ground truth, initialization values, and convergence criterion.

TABLE I
EXPERIMENT RESULTS OF POINT REGISTRATION WITH ECM AND ICP (FIRST GROUP: WITHOUT NOISE)

	ECM	Standard ICP
Correct Match	$\frac{36}{40} = 90\%$	$\frac{15}{40} = 37.5\%$
Registration Accuracy (*mm*)	0.36 (Mean) 0.78 (STD) 2.88 (Max) 0.01 (Min)	56.90 (Mean) 49.32 (STD) 154.80 (Max) 0.01 (Min)
Registration Accuracy (*mm*) (only correct matches)	0.13 (Mean) 0.38 (STD) 1.87 (Max) 0.01 (Min)	0.28 (Mean) 0.58 (STD) 1.77 (Max) 0.01 (Min)
Number of Iterations	28.6 (Mean) 7.0 (STD) 48.0 (Max) 19.0 (Min)	36.0 (Mean) 15.9 (STD) 74.0 (Max) 15.0 (Min)

TABLE II
EXPERIMENT RESULTS OF POINT REGISTRATION WITH ECM AND ICP (SECOND GROUP: WITH GAUSSIAN WHITE NOISE, STANDARD DEVIATION $\sigma = 1$MM)

	ECM	Standard ICP
Correct Match	$\frac{15}{40} = 37.5\%$	$\frac{0}{40} = 0\%$
Registration Accuracy (*mm*)	3.19 (Mean) 2.17 (STD) 11.01 (Max) 0.70 (Min)	76.88 (Mean) 37.15 (STD) 125.11 (Max) 9.19 (Min)
Registration Accuracy (*mm*) (only correct matches)	1.51 (Mean) 0.44 (STD) 2.06 (Max) 0.70 (Min)	-- (Mean) -- (STD) -- (Max) -- (Min)
Number of Iterations	53.4 (Mean) 19.1 (STD) 100.0 (Max) 22.0 (Min)	27.8 (Mean) 14.2 (STD) 69.0 (Max) 12.0 (Min)
Noise Strength (*mm*)	X-axis: 0.80 (Mean), 0.60 (STD), 5.05 (Max), 0.00 (Min) Y-axis: 0.80 (Mean), 0.60 (STD), 5.22 (Max), 0.00 (Min) Z-axis: 0.80 (Mean), 0.60 (STD), 4.64 (Max), 0.00 (Min)	

For each trial on the ECM algorithm, 50 points randomly chosen from 24994 observed points are transformed to serve as model points, and these 50 model points are also used by one of the 40 trials on the ICP algorithm. The observed experiment outputs within this group are the number of iterations and correct matches, and the registration accuracy.

Experimental setups of the group with Gaussian white noise are identical to those of the other group, except that the observation points are disturbed by Gaussian white noise with the mean centered at each observation point and the standard deviation $\sigma = 1$mm. The experiment outputs including the number of correct matches and iterations, the registration accuracy, and the noise strength are observed.

B. Results

The experiment results of the group without noise are listed in Table I.

Although the initialization parameters remain unchanged for all 80 trials in this group as stated above, they can be considered as the randomly chosen values for each trial, since the 50 model points are selected randomly. Initialized by different random parameters, the ECM algorithm could achieve correct matches in 90% of 40 trials, whereas the success ratio of the standard ICP algorithm is less than 40%, as the first row ("Correct Match") of Table I shows. This implies that the ICP method is more sensitive to the initialization than the ECM method.

The results shown in the row of "Registration Accuracy" indicate that the point registration with ECM provides us with more consistent estimations of the transformation parameters than that with ICP. The huge variation of the registration accuracy from ICP can be explained by the fact that the ICP algorithm easily falls into local minima and leads to different registration results with different initializations.

In the third row of Table I, we notice that given an appropriate initialization, the ICP algorithm can reach as high a registration accuracy as the ECM algorithm within the situation without noise.

Furthermore, the last row of Table I demonstrates that the efficiency of the ECM method is comparable with that of the ICP method in the ideal case.

Table II gives the experiment results of the group with Gaussian white noise (standard deviation $\sigma = 1$mm). It illustrates that in case of the presence of noise, ICP works efficiently but not robustly, whereas ECM is not so efficient but quite robust.

IV. CONCLUSION

In this paper, we performed the implementation of an ECM algorithm and a standard ICP method. Moreover, we designed an experiment to register a small number of model points with a huge amount of observation points on a pelvis model. By conducting experiments in scenarios without noise and with Gaussian white noise, we evaluated and compared the performance of both algorithms for point registration.

The experiment results show that both algorithms can obtain an accurate registration, whereas the ICP method requires the appropriate initialization to converge globally. With respect to different initializations, the registration results from ICP are very inconsistent. On the contrary, ECM can produce consistent results, because its CM-steps ensure that it does not become easily captured by local minima. Our experiments also prove that the ECM algorithm is not as sensitive to initial parameter values as the ICP algorithm. We also observed in experiments that ECM is less efficient but more robust than ICP in the case with noise.

In the future, we would like to investigate the extension of the ICP method and the acceleration of the ECM algorithm with e.g. the fast Gauss transform. Based on this work, a further comparison in terms of the computational complexity between ECM and ICP will be covered. Furthermore, both algorithms will be verified and compared in different experiments e.g. in scenarios with outliers. In addition, we are also interested to extend the ECM method to different applications including the non-rigid case.

REFERENCES

[1] H. Hufnagel, X. Pennec, J. Ehrhardt, N. Ayache, and H. Handels, "Generation of a statistical shape model with probabilistic point correspondences and the expectation maximization - iterative closest point algorithm," *International Journal of Computer Assisted Radiology and Surgery*, vol. 2, no. 5, pp. 265–273, 2008. (DOI: 10.1007/s11548-007-0138-9)

[2] C. S. K. Chan, P. J. Edwards, and D. J. Hawkes, "Integration of ultrasound based registration with statistical shape models for computer assisted orthopaedic surgery," In *Proc. SPIE Medical Imaging 2003*, vol. 5032, pp. 414–424, 2003. (DOI: 10.1117/12.480476)

[3] R. Horaud, F. Forbes, M. Yguel, G. Dewaele, and J. Zhang, "Rigid and Articulated Point Registration with Expectation Conditional Maximization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 3, pp. 587–602, 2011. (DOI: 10.1109/TPAMI.2010.94)

[4] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-Squares Fitting of Two 3-D Point Sets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-9, no. 5, pp. 698–700, 1987.