Multi-scale EM-ICP: A Fast and Robust Approach for Surface Registration

Sébastien Granger[1,2] and Xavier Pennec[1]

[1] INRIA, Epidaure Project, Sophia Antipolis, France
[2] AREALL, Neuilly-sur-Seine, France
{Sebastien.Granger, Xavier.Pennec}@sophia.inria.fr

Abstract. We investigate in this article the rigid registration of large sets of points, generally sampled from surfaces. We formulate this problem as a general Maximum-Likelihood (ML) estimation of the transformation and the matches. We show that, in the specific case of a Gaussian noise, it corresponds to the Iterative Closest Point algorithm (ICP) with the Mahalanobis distance.
马氏距离，计算两个样本之家的相似度
Then, considering matches as a hidden variable, we obtain a slightly more complex criterion that can be efficiently solved using Expectation-Maximization (EM) principles. In the case of a Gaussian noise, this new methods corresponds to an ICP with multiple matches weighted by normalized Gaussian weights, giving birth to the EM-ICP acronym of the method.
The variance of the Gaussian noise is a new parameter that can be viewed as a "scale or blurring factor" on our point clouds. We show that EM-ICP robustly aligns the barycenters and inertia moments with a high variance, while it tends toward the accurate ICP for a small variance. Thus, the idea is to use a multi-scale approach using an annealing scheme 退火机制
on this parameter to combine robustness and accuracy. Moreover, we 目的：概率性寻找全局最优
show that at each "scale", the criterion can be efficiently approximated using a simple decimation of one point set, which drastically speeds up the algorithm.
Experiments on real data demonstrate a spectacular improvement of the performances of EM-ICP w.r.t. the standard ICP algorithm in terms of robustness (a factor of 3 to 4) and speed (a factor 10 to 20), with similar performances in precision. Though the multiscale scheme is only justified with EM, it can also be used to improve ICP, in which case the performances reaches then the one of EM when the data are not too noisy.

Keywords: Surface registration, ICP algorithm, EM algorithm, Multi-scale.

总结而言：
1. 使用最大似然估计；
2. 使用EM算法
结果：
1. robustness；2. speed

1 Introduction

This paper describes a new method for the registration of surfaces. This kind of registration is usually performed using one of the multiple variations around the Iterative Closest Point (ICP) algorithm [1,19]. This algorithm is quite fast and

A. Heyden et al. (Eds.): ECCV 2002, LNCS 2353, pp. 418–432, 2002.

accurate in many cases, but, as any method minimising a non-convex cost function, it lacks robustness w.r.t. the initial transformation because of local minima. For surface registration, we experienced that these local minima are quite numerous, probably due to the discrete sampling of the surface. The computation time is also proportional to the number of points, which can be prohibitive when registering two large sets of points.

As our registration method is developed as a component of a per-operative system for surgery guidance namely the *DentalNavigator* system (patent pending) developed by AREALL [5], and based on finely sampled surfaces (containing 30000 to 200000 points), we need to drastically improve both the robustness and the computation times without loosing the accuracy.

We found in the literature three main classes of methods to improve the robustness of ICP. The first class uses robust estimators [19,11] to deal with outliers (i.e. erroneous or occulted points). However, these techniques are not designed to be robust w.r.t. the initial transformation, which is our main point of interest in this paper. The second class is based on stochastic approaches [14]. This kind of variants are quite efficient, but usually require more computation time. The last class is based on the smoothing of the criterion. The most interesting work has been presented by Rangarajan et al., who introduced multiple weighted matches justified by a probabilistic vision of the matching problem. They developed matching models based on Gaussian weight (SoftAssign [16]) and Mutual Information [15], leading to a smaller number of local minima and thus presenting the most convincing improvements.

随机搜索寻找较好初值

这一段的论文还是需要看一下的

The last two classes are usually based on a variance parameter. For high values, the algorithm is more robust but less accurate, while it behaves almost like the ICP for low values. The combination of robustness and accuracy is then obtained through an annealing scheme on this variance parameter. Other works suggest that improvements in terms of speed could be obtained with algorithmic tricks or by decimating the sets of points [12]. Up to our knowledge, none of these works provide guarantees on the convergence of the algorithm nor on the conservation of the robustness and accuracy properties.

鲁棒性和speed不太可以兼得，但是通过退火算法可以解决

Our work was inspired by Rangarajan's probabilistic approach. We first develop in Section 2 a general Maximum-Likelihood (ML) estimation of the transformation and the matches. We show that, in the specific case of a Gaussian noise hypothesis, it corresponds to the Iterative Closest Point algorithm (ICP) with the Mahalanobis distance. Then, considering matches as a hidden variable, we obtain a slightly more complex criterion that can be efficiently solved using Expectation-Maximization (EM) principles. Still with Gaussian noise, this new methods roughly corresponds to an ICP with multiple matches weighted by normalized Gaussian weights, giving birth to the EM-ICP acronym of the method. A similar approach has been independently developed in [2].

Section 3 investigates the influence of a new parameter: the variance of the Gaussian. We show that EM-ICP robustly aligns the barycenters and inertia moments with a high variance, while it tends toward the accurate ICP for a small variance. Thus, we propose to view this variance as a "scale or blurring factor"

on our point clouds. This leads to a multiscale scheme designed to improve robustness without penalizing the accuracy. Due to the high computational cost at large scales, we efficiently approximate the criterion using an original decimation of one point set, which drastically speeds up the algorithm. Finally, we analyze the relation between our method and the multi-scale framework used in standard image (i.e. intensity) analysis registration.

Last but not least, Section 4 presents experimental results showing drastic gains in terms of computation time and robustness.

2 Statistical Derivation of the EM-ICP Algorithm

将sample(采样)视为整个模型的一个测量

In this section, we model the scene as a set of noised measurements of the model points. The priors are the a-priori probability that a given scene point is a measure of a given model point. The two unknowns are the transformation and the matches (i.e., for each scene point, the variable that indicates which model point is measured). A Maximum Likelihood (ML) estimation of these two unknown leads, in the case of a Gaussian noise with uniform priors, to the ICP criterion using the Mahalanobis distance. Since the matches are usually not a variable of interest, we focus then on a ML estimation of the transformation alone. This leads to a new criterion that lacks an efficient optimisation method.

对后验概率进行加权

Going back to the first criterion, another idea is to compute its expectation with respect to every possible matches weighted by the corresponding a-posteriori probability. This efficient optimisation scheme gives an algorithm close to the ICP, but with multiple weighted matches. We show that these two approaches are in fact a particular case of the general EM algorithm, and thus combine an efficient optimization scheme with a well posed criterion with a convergence proof.

2.1 ML Estimation of the Transformation and the Matches

Let s_i be the $n_{\mathcal{S}}$ points of the scene $\mathcal{S}$, m_j the $n_{\mathcal{M}}$ points of the model $\mathcal{M}$ and T a rigid transformation from the scene to the model. Assuming that $T \star s_i$ *is homologous to* m_j (i.e. a measure of m_j) with a known noise model, its density probability function can be defined by: $p(s_i|m_j, T) = p(T \star s_i|m_j)$.

1. Noise model

In the case of an additive and centered Gaussian noise of covariance Σ, this probability can be defined using the Mahalanobis distance $\mu^2(x, y) = (x - y)^t.\Sigma^{(-1)}.(x - y)$: $p(s_i|m_j, T) = k^{-1}.\exp(-\mu^2(T \star s_i, m_j)/2)$.

2. how to define noise model

To represent the matches estimation, we use a binary matrix A: $A_{ij} = 1$ if s_i matches m_j and 0 otherwise. Since each scene point s_i is assumed to correspond exactly to one model point with index say $j^\star$, we have $A_{ij} = \delta_{jj^\star}$ and $\sum_j A_{ij} = 1$ for all scene indices i. In order to represent random matches, we use a random matching matrix $\mathbf{A}$. Each possible matching matrix A has a probability $p(A) = P(\mathbf{A} = A)$ and verifies the previous constraints: $\overline{A_{ij}} = E(\mathbf{A}_{ij}) = P(\mathbf{A}_{ij} = 1) \in [0, 1]$ and $\sum_j \overline{A_{ij}} = 1$. Finally, since scene points will assumed to be independent (see below), and using $\alpha^1 = \alpha$ and $\alpha^0 = 1$, we can write: $p(A) = \prod_{ij/A_{ij}=1} p(\mathbf{A}_{ij} = 1) = \prod_{ij} (\overline{A_{ij}})^{A_{ij}}$.

将Match看作一种随机变量

The first example of such a random matrix is the a-priori probability of the matches: this is the law giving the probability that a given scene point is a measure of a given model point knowing nothing else: $p(\mathbf{A}_{ij} = 1) = \overline{\pi_{ij}}$. A relevant choice is usually the uniform law $\overline{\pi_{ij}} = \frac{1}{N_{\mathcal{M}}}$.

给了一个先验概率的例子（矩阵A的例子）

Now, the joint probability of s_i and $\mathbf{A}_{ij}$ is simply: $p(s_i, A_{ij} = 1|\mathcal{M}, T) = \overline{\pi_{ij}}.p(s_i|m_j, T)$. We can write this last equality for a whole row of A: $p(s_i, A_i|\mathcal{M}, T) = \prod_j (\overline{\pi_{ij}}.p(s_i|m_j, T))^{A_{ij}}$, and assuming that *all scene points are conditionally independent*, the joint probability of the scene and the matches is the product of each of them:

两个事件相互独立

这个就是最大似然估计的公式
从样本估计参数

$$p(\mathcal{S}, A|\mathcal{M}, T) = \prod_i p(s_i, A_i|\mathcal{M}, T) = \prod_{ij} (\overline{\pi_{ij}}.p(s_i|m_j, T))^{A_{ij}} \quad (1)$$

将乘机转化为和的形式

Finally, we are looking for the transformation that maximises the likelihood of the observed scene, or equivalently minimises its negative log:

$$C_{ICP}(T, A) = -\log p(\mathcal{S}, A|\mathcal{M}, T) = \sum_{ij} A_{ij}.(-\log p(s_i|m_j, T) - \log \overline{\pi_{ij}}) \quad (2)$$

For an homogeneous Gaussian noise and uniform priors, this simplifies into:

给定matching是一定的，并且噪声模型的类型是高斯噪声

$$C_{ICP}(T, A) = \frac{1}{2} \sum_{ij} A_{ij}.\mu^2(T \star s_i, m_j) + Cte \quad (3)$$

One recognises here the standard ICP criterion using the Mahalanobis distance. This proves that ICP is no more than a maximum likelihood approach of the registration problem. Moreover, [9] showed that this is the best (minimal variance) estimator. Note that the criterion is invariant w.r.t a global scaling of the noise covariance. This property will not hold any more for the following EM formulation.

2.2 ML Estimation of the Transformation Ignoring the Matches

In the previous section, the transformation and the matches were both estimated by the maximisation of the likelihood of the scene. But for registration purposes, the matches are not parameters of interest. Two different ideas arise: on the one hand, we could maximise the scene likelihood knowing only the transformation. We present this method in this section. On the other hand, there is no reason to consider only the most likely matches, especially when there are ambiguities. We could consider all the possible matches, compute their respective probability, and use the expectation of the previous criterion (Eq. 2) w.r.t. these probabilities (a kind of Bayesian estimation). This is presented in Section 2.3.

感觉有必要补充一下概率论的相关知识，关于概率估计，似然等内容

Let us go back to the probability of each scene point. As we don't want to deal with the matches estimation, we have to consider the likelihood of a scene point knowing only the transformation:

$$p(s_i|\mathcal{M}, T) = \sum_{\{A_i\}} p(s_i, A_i|m_j, T) = \sum_j \overline{\pi_{ij}}.p(s_i|m_j, T)$$

This amounts to seeing the scene points as measurements of a mixture of probabilities around the model points: there is no more homology between a scene point and one of the model points. This interpretation is specially adapted for our case since the model is a surface and not a collection of landmarks. Assuming

once again the independence of the scene points measurements, the likelihood of the scene and its negative log are :

到这里转化为似然估计

$$p(\mathcal{S}|\mathcal{M},T) = \prod_i \sum_j \overline{\pi_{ij}}.p(s_i|m_j,T) \quad (4)$$

$$C_{EM}(T) = \sum_i (-\log \sum_j \overline{\pi_{ij}}.p(s_i|m_j,T)) \quad (5)$$

Unfortunately, this criterion has no closed form solution and we do not know any robust and efficient method to minimise it directly.

2.3 "Bayesian" Estimation of the Transformation

We now turn to the second method, which implies the computation of the a-posteriori probability of each possible matches matrix A knowing the scene, the model and the transformation. This can be achieved using Bayes rule, and Eq. 4 and 1. To simplify the notations, we denote by $\mathbf{A}_T$ the random matching matrix defined by:

$$p(\mathbf{A}_T = A) = p(A|\mathcal{S},\mathcal{M},T) = \frac{p(\mathcal{S},A|\mathcal{M},T)}{p(\mathcal{S}|\mathcal{M},T)} = \prod_{ij} \left(\frac{\overline{\pi_{ij}}.p(s_i|m_j,T)}{\sum_k \overline{\pi_{ik}}.p(s_i|m_k,T)} \right)^{A_{ij}} \quad (6)$$

Since we have $p(A_T = A) = \prod_{ij} (\overline{(A_T)_{ij}})^{A_{ij}}$, we obtain by identification:

$$\overline{(A_T)_{ij}} = \frac{\overline{\pi_{ij}}.p(s_i|m_j,T)}{\sum_k \overline{\pi_{ik}}.p(s_i|m_k,T)} \quad (7)$$

Thus, the idea is to optimise the expectation of our first criterion (Eq. 2) w.r.t. this law:

$$C_{\mathbf{A}_T}(T) = \mathbf{E}_{\mathbf{A}_T}[C_{ICP}(T,A)] = \sum_{ij} \overline{(A_T)_{ij}}.(-\log p(s_i|m_j,T) - \log \overline{\pi_{ij}}) \quad (8)$$

At this point, the algorithm is obvious: starting from a first estimation of the transformation, we alternatively compute the probability of matches $\overline{(A_T)_{ij}}$ and optimise this new criterion. Unfortunately, this new algorithm does not correspond to the minimisation of a well posed criterion, as $\overline{(A_T)_{ij}}$ depends on T and vice versa. Thus, the convergence is not ensured.

2.4 Equivalence of the Two Approaches: The EM Algorithm

In fact, it turns out that the algorithm justified in Sec. 2.3 is no more than the optimisation of the criterion presented in Sec. 2.2, through an Expectation-Maximisation (EM) approach [3,13]. Thus, the algorithm is ensured to converge.

To relate the criterion $C_{EM}(T) = -log(p(\mathcal{S}|\mathcal{M},T))$ with the optimisation method, let us introduce the matching matrix in the criterion using Bayes rule and take its expectation for any law on A: $C_{EM}(T) = E_{\mathbf{A}}(C_{EM}(T)) = -E_{\mathbf{A}}(\log p(\mathcal{S},A|\mathcal{M},T)) + E_{\mathbf{A}}(\log p(A|\mathcal{S},\mathcal{M},T))$ Since this criterion is still independent of $\mathbf{A}$, we create a secondary criterion that depends explicitly on them by adding the Kullback-Leibler distance between $\mathbf{A}$ and the a-posteriori law on the matches: $KL(\mathbf{A}||\mathbf{A}_T) = E_{\mathbf{A}}(\log p(A) - \log p(A|\mathcal{S},\mathcal{M},T))$. This distance is positive and null only for $\mathbf{A} = \mathbf{A}_T$:

$$C_{EM}(T,\mathbf{A}) = C_{EM}(T) + KL(\mathbf{A}||\mathbf{A}_T) \quad (9)$$

$$= E_{\mathbf{A}}(\log p(\mathcal{S},A|\mathcal{M},T)) + E_{\mathbf{A}}(\log p(A)) \quad (10)$$

The EM algorithm is the alternative minimisation of this criterion.

E-Step: estimation of the matches. Here, T is fixed and we minimise the criterion w.r.t. $\mathbf{A}$. Using Eq. (9), we are left with $\mathbf{A} = \mathbf{A}_T$. Since the distribution of the $(\mathbf{A}_T)_{ij}$ is binary and normalized, it can be uniquely characterized by its mean value $(A_T)_{ij}$. As scene points are independant, they can be treated seperatly. In practice, for each scene point, we search all close model points, and compute their weights using Eq. (7).

M-Step: estimation of the transformation. Now, the matches probability are fixed and equal to $\mathbf{A}_T$, and we minimise the criterion w.r.t. T. Using Eq. (10) and dropping the constant term $E_{\mathbf{A}}(\log p(A))$, we are left with the minimisation of the expectation of $\log(p(\mathcal{S}, A|\mathcal{M}, T)))$, which is exactly the criterion (8).

Optimality and convergence: After the E-Step, we have exactly $KL(\mathbf{A}||\mathbf{A}_T) = 0$ and then $C_{EM}(T, \mathbf{A}) = C_{EM}(T)$. This proves that our "Bayesian" estimation is a computational trick for minimising $C_{EM}(T)$ and thus the algorithm is ensured to converge toward a local minimum of $C_{EM}(T)$.

3 Multiscale Gaussian EM-ICP with Decimation

We focus in this section on the simple case where the noise is homogeneous, Gaussian and isotropic with uniform priors on the matches. In this case, we are left with only one free parameter: the variance of the Gaussian noise. We first analyze the influence of the variance on the criterion shape and show that it behaves as a smoothing factor. This leads to a multiscale scheme designed to improve robustness without penalizing the accuracy: due to the high computational cost at coarse scales, we combine an annealing scheme with a simple but very efficient decimation technique as an approximation of the criterion. This technique is very similar to the multi-scale approaches used in standard image (i.e. intensity) analysis domains, and lead to substantial improvements in terms of robustness and computation times.

3.1 EM-ICP for a Homogeneous Gaussian Noise and Uniform Priors

For an isotropic and homogeneous Gaussian noise, the Mahalanobis distance reduces to a rescaled distance: $\mu^2(T \star s_i, m_j) = \frac{1}{\sigma^2}\|T \star s_i, m_j\|^2$, and the two EM steps turn out to be particularly simple:

$$\overline{(A_T)_{ij}} = \frac{\exp(-\|T \star s_i - m_j\|^2/2\sigma^2)}{\sum_k \exp(-\|T \star s_i - m_k\|^2/2\sigma^2)} \tag{11}$$

$$C_{\mathbf{A}_T}(T) = n_{\mathcal{S}}.dim.\log(\sigma) + \frac{1}{2\sigma^2}\sum_{ij}\overline{(A_T)_{ij}}.\|T \star s_i - m_j\|^2 \tag{12}$$

Practically, the only difference with the ICP for optimizing the transformation is the presence of non-binary weights. In the rigid case, this can be solved by a straightforward adaptation of the SVD or the unit quaternion methods [4].

One can even simplify Eq. (12) by considering that each scene point s_i contribute to the criterion through $\sum_j \overline{(A_T)_{ij}}\|T \star s_i - m_j\|^2$. As the weights are

normalized, this is equal to $\|T \star s_i - \overline{m_i}\|^2 + \sum_j \overline{(A_T)_{ij}} \|\overline{m_i} - m_j\|^2$, where $\overline{m_i}$ is the barycenter of the m_j weighted by the $\overline{(A_T)_{ij}}$. As the second term is constant during the M-Step, we are left with the optimization of $\sum_i \|T \star s_i, \overline{m_i}\|^2$, which is a standard least-square.

One important practical problem is the rejection of outliers. The theoretical developments of the previous section could easily be adapted to add a probability to match a scene point to the background, as done in [18, pp.78]. In our case, this turns out to be almost equivalent to thresholding the Mahalanobis distance. Thus, we only need to look for matches up to a maximum distance (e.g. $3.\sigma$). Model points farther away are considered as outliers and are given a null weight.

To determine the variance, we may use a ML approach, which amounts to minimizing the EM criterion with respect to σ given a transformation and a set of (probabilistic) matches. This gives a simple RMS estimation:

$$\sigma^2 = \frac{1}{n_{\mathcal{S}}.dim} \sum_{ij} \overline{(A_T)_{ij}}.\|T \star s_i, m_j\|^2. \tag{13}$$

One could think to estimate the variance this way at each EM iteration as a third step of the optimization process, but we experimentally observed that the decrease was too sharp and was sticking the algorithm in local minima.

3.2 Criterion Shape and Annealing Scheme

To understand the influence of the variance, we analyzed in [7] the asymptotic values of the criterion and showed that, for high values of the variance, it simply aligns inertia centers and moments of the two point sets, while it reaches the standard ICP criterion for small values of σ. Intermediate values of the variance are presented in Fig. 1: one clearly see that the brittle shape of the ICP criterion (with numerous local minima) is smoothed into an almost quadratic shape for a higher variance with a decreasing number of local minima for intermediate values. One can also see that the correct (i.e. the global) minimum is unfortunately shifted when the variance is increased. This suggests that we should start from a large variance, which should guaranty the robustness, and track this minimum as the variance decrease to end up with the real noise variance that will ensure the most accurate results.

Since using the ML estimation of the variance imposes too fast a decay, we chose to impose a slight but regular decrease by dividing the variance parameter after each iteration with a fixed value called annealing coefficient (typically 1.1). Notice that doing so can sometimes lead to an increase of the criterion value between two step. An alternative approach would be to wait for the convergence before decreasing the variance parameter, but we experimented that it was slower and less efficient. The annealing process is stopped either when the variance reaches a predefined value, computed on a good registration of typical data-sets, or when it reaches the value estimated using Eq. 13 at the current iteration. The initial value of the variance obviously depends on the quality of the initialization (i.e. the initial estimation of the transformation). We have no theoretical way of evaluating this value yet. In our experiments, we used a value varying from 16 to 100 times the typical real noise variance.

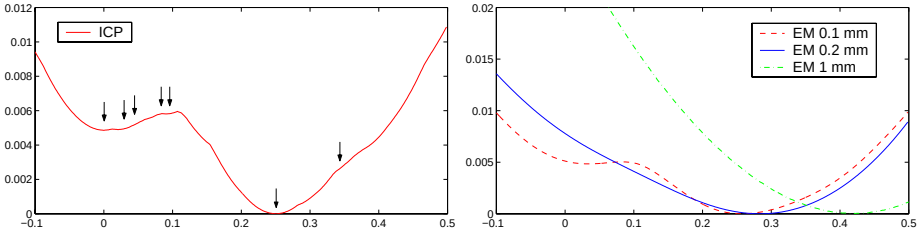


Fig. 1. *Criterion vs Z-translation for ICP (on the left) and for EM (on the right).* There are two relevant minima for ICP (0, where the algorithm stopped, and 0.25, where criterion is minimum) and irrelevant ones (e.g. 0.1 and 0.3). The EM criterion has a much smoother shape. For an under-estimated variance (0.1 mm) the irrelevant minima have disappeared. For the approximative noise variance (0.2 mm), even the first minima has disappeared. For a larger (overestimated) variance (1 mm), the criterion is almost quadratic, but the global minimum has been shifted.

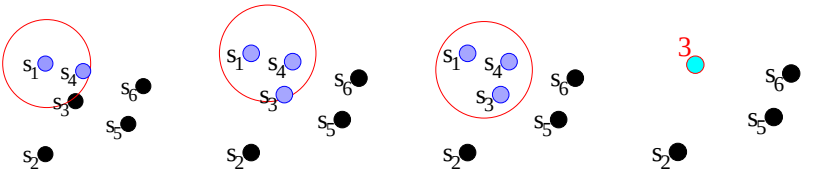
3.3 Decimation

In preliminary experiments [7], we dealt with the case of a small number of points (below a hundred) measured on a surface. EM-ICP appeared to be more accurate than ICP, and the annealing scheme was affordable enough to provide a high robustness. In this paper, we are interested in large number of scene points (typically 20000 to 50000). With the ICP algorithm, most of the computation time is spent in the closest point search. This is usually implemented using a kD-tree or some similar method and the search is limited to the threshold distance defined for outliers rejection. The same technique is of course applied when using the EM-ICP algorithm, except that all points closer than a maximal Mahalanobis distance are taken into account. one-to-all 的方式也可以使用outlier rejection;

With the annealing scheme, the problem is that the maximum Euclidean search distance is then proportional to the standard deviation: with a large variance, virtually all model points have to be retrieved and the kD-tree becomes completely inefficient. This has an "exponential" influence on the computation time that practically forbid the use of the annealing scheme. The main idea was then to see the variance as a scale parameter: at a coarse scale, we do not need to be accurate and we can approximate roughly our data-set. When the variance goes down, we have to refine our approximation to increase the accuracy. 这里的搜索仍然是按照 kd-tree, 但是是 one-to-many

To speed-up the algorithm at a given scale, the basic idea is to observe that if n scene points are close enough (w.r.t. the variance), they will share almost the same closest points within the model with almost the same weights. Thus, approximating these n points by considering n time one of them will optimize a good approximation of our criterion (n is called the *decimation weight*). In fact, it turns out that we have to use their barycenter to obtain a second order approximation of the criterion (detailed proof in [6]). This leads to a simple decimation technique that merges scene points that are close to each other. More precisely, we consider that the distance of a point to its approximation has to be inferior to $\alpha.\sigma$.

In fact, any decimation technique that replace a subset of close points by their barycenter will be compatible with EM-ICP. One could imagine lots of



(a) Sphere init. (b) Iteration 1 (c) Iteration 2 (d) Fusion

Fig. 2. The greedy sphere decimation algorithm. At the beginning, we define a *remaining list* containing all the points of the original cloud and we create an empty *decimated cloud*. We iteratively add points in the decimated cloud as follows. Starting from a remaining point, we initialize a sphere of radius $\alpha.\sigma$ centered on this point (a). Then, we iteratively find all remaining points in this sphere, compute their barycenter $\overline{s}$, and move the center of the sphere to this barycenter (b,c). After convergence, we create a new point in the "decimated cloud" at the center of the sphere with a weight equal to the number of remaining points within the sphere, and we remove these remaining points (d). The whole process ends up when there are no more remaining points.

different ways of finding these subsets of points. We present in Fig. 2 a simple and very fast technique: the sphere decimation. It consists in adjusting a sphere of radius $\alpha.\sigma$ centered on the barycenter of the points it contains. The modifications of the EM-ICP algorithm including the decimation and annealing schemes are presented in Table 1.

Table 1. Pseudo-code of the Multiscale EM-ICP registration algorithm.

Initialization : Compute a first estimation of T, and set σ^2 to its initial value.
Repeat
  Decimation : Decimate the scene with a sphere radius $\alpha.\sigma$.
  E-Step : For each s_i in the decimated scene with decimation weight n_{s_i}:
    Search all m_j such that $\underline{\|T \star s_i - m_j\|^2} < \sigma^2.\mu^2_{max}$ using a kD-tree
    Compute the weights $\overline{(A_T)_{ij}}$ using Eq. 11
  M-Step : Re-estimate T by minimising $\sum_{ij} n_{s_i} \overline{(A_T)_{ij}}.\|T \star s_i - m_j\|^2$.
  Annealing : Divide σ^2 by the annealing coefficient. If σ^2 is below the final noise variance, set it to the final noise variance.
Until convergence

3.4 Analogy with Multi-scale Intensity-Based Techniques

As the variance can be considered as a scale parameter, it is interesting to determine how our annealing and decimation scheme relates to multiscale techniques already developed for intensity-based registration. We analyze in this section the coarse-to-fine aspects, realized by the coupling of the EM algorithm and the annealing scheme, and the down-sampling aspects, realized by the decimation, and linked to the multi-grid approach with intensity images. Our discussion will be quite pragmatic, though we think it could be integrated into a rigorous multi-scale framework (see for instance [10]).

Coarse-to-fine approaches. The main idea underlying multiscale intensity- (i.e. image-) based registration approaches is to get rid of small details that create local minima and trap the algorithm when the initialization is too far away from the solution. This is especially the case for non-rigid registration [17]. This is usually implemented using a pyramid of blurred images with an increasing Gaussian kernel [8].

Using a Gaussian mixture model for representing the probability of measuring points follows essentially the same idea: we are representing the pdf of measuring one point of the model as a blurred version of the original (Dirac) model measurements. As the variance grows to larger values, we will loose the precise local shape of the surface to focus on a more global shape.

The main difference with images is that our measurements are not on a regular grid: precomputing the pdf on a regular grid would forbid the use of an efficient ICP-like algorithm. This is why the blurring is introduces directly within the algorithm via the EM weights. Moreover, the normalization has to be done for each scene point independently (as distances are different) and cannot be incorporated into the blurring kernel. This could be a difficult point in the rigorous modeling of our algorithm as a multiscale approach.

In the multiscale theory, one could show that the causality hypothesis implies that the local minima at a fine scale are gradually merging with the scale, and that there is no bifurcation (creation of new minima). Although we experimented a similar behavior with EM-ICP (see Fig. 1), we lack a rigorous multiscale setting to prove it. Such a result would be especially interesting since we know that there is a unique minimum at the largest scale (if the inertia moments of the point clouds are sufficiently anisotropic): could we find a path along scales leading from this unique minimum to the global minimum at the finest scale ? If there exists moreover a technique to choose the right path, we could obtain a perfectly robust algorithm.

Decimation. Multi-scale techniques are often coupled with a down-sampling of the image at each-scale, designed to improve the computational efficiency. This down-sampling is usually integrated into a multi-grid approach where the resolution of images is for instance divided by two at each coarser scale. This is implemented through a resampling algorithm that takes into account the Gaussian blurring, and is designed to minimize the impact on the registration criterion while reducing drastically the volume of data.

Our decimation technique was designed with the same concern, but there are some major differences that could have a theoretical impact. Firstly, the scale does not need to be divided by two at each scale to obtain an efficient algorithm: we can have a smoother scale evolution that gives better results (see Section 3.2). The counterpart is that we cannot construct a pyramid using a recursive technique and we have to down-sample the set of points at each scale from the original set of points. In our case, this is not a problem as this step is computationally cheap. Secondly, our decimation is applied before the implicit Gaussian blurring, and not concurrently as for images. This means that we compute an ad-hoc approximation of the point location pdf, and not the best

possible one. However, the possible gain of such an optimisation would probably be discarded by the extra computation time.

4 Experiments

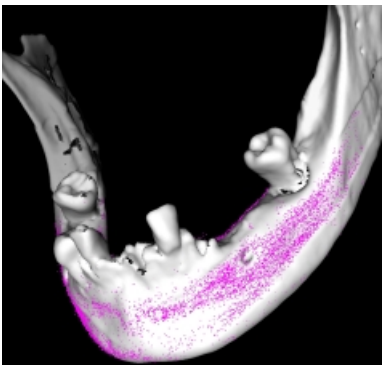
We evaluate in this section the comparative robustness, computation time and repeatability of the ICP and EM algorithms with respect to the decimation and annealing parameters. We present the computation time of the overall registration process (on a PC Pentium III 1Ghz using linux), thus taking into account each iteration time (including decimation) and the number of iterations.

To evaluate the robustness w.r.t. the initial transformation, we performed between about 1000 and 5000 registrations with the same parameters and various initial transformations obtained by composing different translations with a "ground truth" transformation (a visually checked result considered as the reference). These translations are defined on a regular grid going from -8 to +8 mm in each directions. Here, we did not applied rotations since we previously experimented that the algorithm was more robust w.r.t. rotations than translations. The robustness is then evaluated by the percentage of correct convergence.

In fact, correct convergences does not always exactly correspond to the global minimum, because there can be other local minima in its immediate vicinity. This introduces an intrinsic variability of the results that we call *internal error* or repeatability. This error can be interpreted as the consequence of small errors in the matches estimation. It is directly measured by the variability of the correct transformations obtained in the robustness experiments, and we present it as a "Target Registration Error" (the RMS of points induced by the transformation error) in the target area, in our case the whole jaw.

This internal error, due to the variability of the initial transformation, is only a part of the registration error. Extra errors are introduced by the noise on the data and the fact that the scene points are not really independent nor homologous to the model points, as supposed in section 2.1. To measure the *global error*, we are currently preparing controlled experiments on real data.

4.1 Heterogeneous But Precise Data



In this experimental setup, we have two sets of points measured on the surface of a dry (ex-vivo) jaw. The model (in white) is the surface of the jaw segmented from a CT-Scan of resolution 0.25*0.25*0.5 mm. This is a regularly sampled surface formed by 133027 triangulated points. The scene (purple points) is obtained via a sensor mounted on a passive robotic arm with an accuracy of about 0.05 mm. This set of 25000 points is very heterogeneous, and presents many packets of highly correlated points.

We used the following default parameters: a standard-deviation of the real noise of $\sigma_{final} = 0.3$ mm, a maximum Mahalanobis distance of $\mu^2_{max} = 9$, an initial variance coefficient of $a_{init} = 16$ (and thus the initial standard-deviation is $0.3 * \sqrt{16} = 1.2$ mm), an annealing coefficient of $c = 1.1$. By default, the weights of the decimated scene points were not used (see *Decimation weights* below). In the following table, we summarized the robustness, computation time (in seconds) and repeatability (in mm) results with respect to the modified parameters.

	ICP			EM		
	% correct	Time	Int. error	% correct	Time	Int. error
No decim., no annealing	25	320	< 0.01	27	530	< 0.01
No decim.	33	360	< 0.01	30	613	< 0.01
Decim. at σ	58	90	< 0.01	55	155	< 0.01
" with decim. weights	36	75	< 0.01	34	145	< 0.01
Decim. at 2σ	59	27	0.03	53	48	< 0.01
Decim. at 2σ, $a_{init} = 36$	80	25	0.03	71	43	< 0.01
Decim. at 2σ, $a_{init} = 100$	92	25	0.03	88	58	< 0.01
Decim. at 4σ	49	7	0.03	53	13	< 0.01
Decim. at 8σ	27	2	0.24	37	4	< 0.01

Computation time: The EM algorithm is only two times slower than the ICP algorithm, and the decimation speeds up both algorithm by the same factor (from 5 to 50!). Interestingly, the computation time is quite constant when the initial scale grows up. Without decimation, one would have observed at least a quadratic increase.

Decimation weights: as we are dealing with non-homogeneous and correlated measurements, we believe that forgetting the *decimation* weights (though it is theoretically incorrect) tends to produce a more uniform repartition of uncorrelated scene points. The robustness results provides an experimental evidence of this effect.

Size of the decimation spheres: surprisingly, the decimation induces an increase of the robustness up to 2σ. This shows that the changes introduced by the decimation are only smoothing the criterion up to this value (and probably de-correlating the measurements). For higher values, the robustness drops off, as expected.

Initial scale value: the robustness is spectacularly increased by starting at a high scale, at a much lower computational cost, thanks to the decimation. Thus, it seems that the annealing scheme is the most important improvement from a robustness point of view. However, the question remains of how to find, from theoretical considerations, a more optimal way of decreasing the variance.

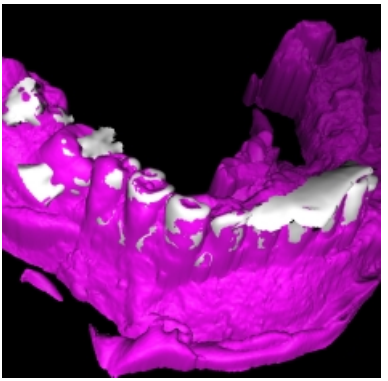
Internal accuracy: the global minimum seems to be very well defined on these data thanks to their very high quality. However, the internal error increases with the size of the decimation sphere for ICP, whereas it is always negligible with the EM algorithm.

The most surprising result on these data is that, except for large decimation, the performances of ICP and EM are equivalent, though the decimation and annealing schemes are not justified with ICP. This is also inconsistent with pre-

vious experiments on small sets of independent points where the EM algorithm was better [7]. We believe that the high precision of the current data points and their important correlation (both criterion assume the independence of scene points) combined with the important decimation and annealing improvements hide here the possible advantages of EM over ICP.

We justified in Sec. 3.3 the decimation of the scene points as realizing an approximation of the criterion. One could easily extend this justification to the decimation of the model point set. We run a few experiments with a comparable decimation radius (between σ and 2σ) which showed that it could lead to an extra gain in terms of computation time, while preserving other performances. However, the main problem is that the kD-tree (used to locate the nearby model points) has to be rebuilt every time the model points are decimated, so that we only gained a factor two on the computation times.

4.2 Homogeneous But Noised Data



Here, the sets of points are measured on the surface of a patient teeth. The model (in white) is the surface of the teeth and the bone segmented from a CT-Scan (same resolution). This is a regularly sampled surface formed by 40897 triangulated points. The Molar and the upper face of incisors are corrupted with artifacts due to fillings. The scene (in purple) is a range image at a resolution of 0.2 mm of a plaster cast of the patient teeth and gum. Notice that only the teeth are visible in both surfaces as the gum of the range image hides the bone surface extracted from the CT.

For this experiment, we used exactly the same default parameters as in the previous section, except that the real noise standard deviation is a bit higher: we determined here that $\sigma_{final} = 0.4$ mm.

	ICP		EM	
	% correct	Time	% correct	Time
No decim., no annealing	47	564	51	800
No decim.	62	579	73	821
Decimation at σ	80	182	91	280
Decimation at 2σ	75	76	88	103
" with decim. weights	45	60	72	98
" with $a_{init} = 36$	94	54	98	65
" with $a_{init} = 100$	100	43	100	64
Decimation at 4σ	47	20	70	35

These results basically confirm the conclusions of the previous experiment, except that EM clearly appears to be more robust than ICP. These better performances are in accordance with our preliminary experiments [7], and tends to

show that EM performs better than ICP in the presence of noise. This effect will have to be confirmed on future experiments. The scene (a range image scanned from above) is less inhomogeneous than in the previous experiments, and the influence of the decimation weights is smaller. This tends to confirm that the gain in robustness previously obtained by the decimation and the suppression of its theoretical weights is specifically due to the non-homogeneity of the these data.

5 Conclusion

We show in this article that the ICP algorithm is a particular case of a more general maximum likelihood approach to the surface registration problem. Then, applying the EM principle, we derive a new variant named EM-ICP. In the case specific case of a Gaussian noise, it amounts to considering a mixture of Gaussians. This introduces a new variable (the variance of the Gaussian) that we interpret as a scale parameter. At a high scale, EM-ICP is very robust as it is almost quadratic, while it behaves like an ICP at a fine scale. We propose to exploit this property in a coarse-to-fine approach based on an annealing scheme to avoid local minima while reaching an optimal accuracy. To bypass the computation time explosion at coarse scales, we design an original approximation of the criterion that may be implemented with a very simple and efficient decimation of the point clouds. The analogy between our method and the multi-scale and multi-grid approaches used in intensity-based image analysis would be interesting to develop as it may lead to a more formal multiscale theory on point sets and surfaces.

Experiments on real data acquired on the jaw and the teeth show a spectacular improvement of the performances of EM-ICP w.r.t. the standard ICP algorithm in terms of robustness (a factor of 3 to 4) and speed (a factor 10 to 20), with similar performances in precision. A closer look at the results show that the main improvements were due to the combination of the annealing and decimation schemes. Though these schemes were only justified with EM, they could also be used to improve ICP, which performances reaches then the one of EM when the data are not too noisy. We also designed a pragmatic variant of the decimation that appears to be very well suited to the case of non-homogeneous and correlated measurements.

We are currently designing controlled experiments on real data to evaluate the global accuracy of the method. As EM presents a better repeatability than ICP, we expect that it will also exhibit a better accuracy. Future work will also focus on the on-line prediction of this accuracy, and the study of the influence of the surface sampling. Another important research axis will be to predict where we should take new measurements in order to improve the registration. This is especially important in view of a real-time per-operative system, in order to guide and speed-up the acquisition of points by the surgeon in an optimal way. From a more theoretical point of view, there is a need for a rigorous scale-space theory on point sets, which could help us to choose the optimal variance decrease in EM-ICP. Last but not least, we believe that this new algorithm could

be adapted to many other applications in the domain of surface and range image registration.

Acknowledgments. This work was supported by a CIFFRE Ph.D fellowship from AREALL, who also provided all the data used in this study.

References

1. P.J. Besl and N.D. McKay. A method for registration of 3D shapes. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992.
2. H. Chui and A. Rangarajan. A feature registration framework using mixture models. In *Proc. MMBIA'2000*, pages 190–197, 2000.
3. C. Couvreur. The EM algorithm: A guided tour. In *Proc. 2d IEEE European Workshop on Computationaly Intensive Methods in Control and Signal Processing (CMP'96)*, pages 115–120, Pragues, Czech Republik, August 1996.
4. D.W. Eggert, A. Lorusso, and R.B. Fisher. Estimating 3-D rigid body transformations : A comparison of four major algorithms. *Machine Vision and Applications*, 9:272–290, 1997.
5. D. Etienne *et al.* A new approach for dental implant aided surgery. a pilot evaluation. In *Proc. CARS'2000*, pages 927–931, 2000.
6. S. Granger and X Pennec. Multi-scale EM-ICP : A fast and robust approach for surface registration. Internal research report, INRIA, 2002.
7. S. Granger, X. Pennec, and A. Roche. Rigid point-surface registration using oriented points and an EM variant of ICP for computer guided oral implantology. Research report RR-4169, INRIA, 2001.
8. J-M. Jolion. *A Pyramid Framework for Early Vision*. Kluwer Academic, 1994.
9. K. Kanatani. *Statistical Optimization for Geometric Computation : Theory and Practice*. Elsevier Science (Amsterdam), 1996.
10. T. Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer academic, 1994.
11. T. Masuda, K. Sakaue, and N. Yokoya. Registration and integration of multiple range images for 3D model construction. In *Proc. ICPR'96*, pages 879–883, 1996.
12. T. Masuda and N. Yokoya. A robust method for registration and segmentation of multiple range images. *Comp. Vision & Image Underst.*, 61(3):295–307, 1995.
13. R.M. Neal and G.E. Hinton. A view of the EM algorithm that justifies incremental, sparse, and other variants. *Learning in Graphical Models*, 1998.
14. G.P. Penney, P.J. Edwards, A.P. King, J.M. Blackall, P.G. Batchelor, and D.J Hawkes. A stochastic iterative closest point algorithm (stochastICP). In Springer, editor, *Proc. of MICCAI'01*, volume LNCS 2208, pages 762–769, 2001.
15. A. Rangarajan, H. Chui, and J.S. Duncan. Rigid point feature registration using mutual information. *Medical Image Analysis*, 3(4):425–440, 1999.
16. A. Rangarajan, H. Chui, E. Mjolsness, S. Pappu, L. Davachi, P. Goldman-Rakic, and J. Duncan. A robust point-matching algorithm for autoradiograph alignment. *Medical Image Analysis*, 1(4):379–398, 1997.
17. P. Thévenaz, U.E. Ruttimann, and M. Unser. A pyramid approach to subpixel registration based on intensity. *IEEE Trans. on Images Proc.*, 7(1):27–41, 1998.
18. W. Wells. Statistical approaches to feature-based object recognition. *Int. Journal of Computer Vision*, 21(1):63–98, 1997.
19. Z. Zhang. Iterative point matching for registration of free-form surfaces. *Int. Journal of Computer Vision*, 13(2):119–152, 1994.