31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Simultaneous reconstruction and registration algorithm for limited view transmission tomography using a multiple cluster approximation to the joint histogram with an anatomical prior

Dominique Van de Sompel and Sir Michael Brady

Abstract— **We develop a novel simultaneous reconstruction and registration algorithm for limited view transmission tomography. We derive a cost function using Bayesian probability theory, and propose a similarity metric based on the explicit modeling of the joint histogram as a sum of bivariate clusters. The resulting algorithm shows a robust mitigation of the data insufficiency problem in limited view tomography. To our knowledge, our work represents the first attempt to incorporate non-registered, multimodal anatomical priors into limited view transmission tomography by using joint histogram based similarity measures.**

I. INTRODUCTION

Limited view transmission tomography is used in a myriad of industrial as well as clinical applications [5]. It is commonly motivated by geometric constraints and limitations on acquisition time and/or patient radiation dose. However, limited view transmission tomography suffers from the limitation that its reconstructions are fundamentally underdefined. This can be understood in terms of the Fourier Slice Theorem [6], which reveals that large swathes of the object's Fourier space remain unmeasured. In this work, we estimate the unsampled information by incorporating an anatomical prior into the reconstruction process.

The use of anatomical priors has been considered previously in emission as well as transmission tomography, where the majority of studies has focused on intensity difference based similarity metrics for monomodal regularization. Examples include the incorporation of planning CTs to regularize intraoperative tomosynthesis reconstructions [1] in transmission tomography, and the simulation of template PET volumes from CT or MRI priors [7] in emission tomography. In these studies it was assumed that the anatomical prior and the object to be reconstructed were aligned *a priori*. Only a few studies have investigated the use of information theoretic similarity measures for multimodal regularization such as mutual information [10] and joint entropy [9], [11], and this only in the field of emission tomography. Mutual information was considered first due to its success in image registration, but it was later demonstrated by Nuyts [9] that joint entropy introduces less bias into the reconstruction and may therefore be more appropriate. In our previous work [13], we built on Nuyts' results and applied joint entropy (JE) regularization to limited view transmisson tomography with *a priori* registered anatomical priors. To our knowledge, this was the first study to do so. In [13], we also identified JE's vulnerability to local optima when used in limited view tomography and proposed two approximate schemes to increase robustness. The first was to minimize the joint entropy of a single bivariate Gaussian fit to the joint histogram during the reconstruction process; the second was to explicitly model and optimize the joint histogram as the sum of a limited number of bivariate probability density distributions. The performance of the first approach was explored in [13] itself, and that of the second in [12], in both cases for pre-registered priors. Since then, we have extended the first approach to the case of non-registered priors and explored its performance in [14]. In the current work, we extend the second approach, i.e. multi-cluster modeling, to the case of non-registered priors as well, and propose a simultaneous reconstruction and registration (SRR) algorithm that can accommodate monomodal as well as multimodal anatomical priors.

D. Van de Sompel is a DPhil candidate in the Department of Engineering Science, University of Oxford, Oxford OX1 3PJ, UK. dominique@robots.ox.ac.uk

M. Brady is with the Department of Engineering Science, University of Oxford, Oxford OX1 3PJ, UK. jmb@robots.ox.ac.uk

II. METHODS

Our experiments in [13] revealed that the JE prior is sensitive to local optima when applied to limited view transmission tomography. In the space of the joint histogram, trapping in local optima amounted to a breaking up of the joint histogram into many small clusters. However, if it is known *a priori* that the ground truth joint histogram contains only a limited number of clusters, the problem can be mitigated by explicitly modelling it as such. This has the effect of reducing the possible number of local optima in the cost function. Here we extend the framework to non-registered priors, and illustrate the method using bivariate Gaussian clusters.

A. Cost function

Our objective is to optimize the joint posterior probability $P(x, p, \theta, \delta | r, B)$, where x is the attenuation map, p is the hidden Markov measure field proposed by Marroquin [8], θ is a vector containing the parameters of each bivariate Gaussian, δ is the deformation field, r is the observed projection data (photon counts), and B is the anatomical prior. The p field is essentially a vector field giving the probability of each pixel x_j belonging to the cluster k. The deformation field δ is a dense vector field defined at every pixel x_j, and gives the displacement from each x_j to the corresponding point in B. The coordinates of the pixels x_j

and B_j are defined in a common reference frame, and are stored as a dense vector field ρ. Hence the intensity of the prior B at the point corresponding to the pixel x_j is given by $B(\rho_j + \delta_j)$. In the remainder of this paper we use the shorthand notation $B_j^* = B(\rho_j + \delta_j)$, where the star is used to differentiate the interpolated intensity $B(\rho_j + \delta_j)$ from the pixel intensity $B_j = B(\rho_j)$. The interpolation of the prior B can be achieved using many different methods, such as linear, quadratic, cubic, or B-spline interpolation. In this study, we use quadratic interpolation as a compromise between accuracy and ease of implementation.

Using Bayes' rule and the definition of conditional probability, the posterior probability can be decomposed as

$$P(x, p, \theta, \delta | r, B) =$$
$$\frac{1}{Z} P(r|x) P(B|x, p, \theta, \delta) P(x) P(p) P(\theta) P(\delta) \quad (1)$$

where we have assumed a uniform distribution for the projection data r, and $P(r|B, x, p, \theta, \delta) = P(r|x)$ since the projection data depends only on the object's attenuation map.

The data likelihood term $P(r|x)$ is modeled by the Poisson distribution, and is given by $\log P(r|x) = \sum_{i=1}^{M} h_i(l_i) + \text{constant}$, where each $h_i(l_i)$ is a concave function. In transmission tomography, they are given by $h_i(l_i) = -\left(r_0 e^{-l_i} + b_i\right) + r_i \log\left(r_0 e^{-l_i} + b_i\right)$ where $l_i = \sum_{j=1}^{N} a_{ij} x_j$. The integer N is the number of pixels in the discretized attenuation map x, a_{ij} is the length of traversal of the i^{th} ray through the j^{th} pixel, r_i is the photon count observed by the i^{th} detector, $r_{0,i}$ is the number of photons leaving the source for the i^{th} ray, and b_i accounts for scatter events.

The remaining terms in Eqn. 1 are independent of the projection data, pertaining solely to the registration of x and B, and the modeling of x in the image as well as joint histogram space. Assuming conditional independence of the pixels, we can express the conditional probability $P(B|x, p, \theta, \delta)$ as

$$P(B|x, p, \theta, \delta) = \prod_{j=1}^{N} \sum_{k=1}^{K} P(B_j^* | f_j = k, x_j, \theta_k, \delta_j) p_{j,k} \quad (2)$$

where K is the number of postulated clusters in the joint histogram, and the cluster labels f have been marginalized out as described by Marroquin [8]. The k^{th} entry of the K-dimensional vector p_j gives the marginal probability that the j^{th} pixel of x belongs to the k^{th} cluster. As noted by Marroquin, the hidden measure field has the advantage that superfluous clusters k are automatically eliminated, resulting in a robustness to overdeclaration of the number K. Assuming bivariate Gaussian distributions for each cluster, we have

$$P(B_j^* | f_j = k, x_j, \theta_k, \delta_j) = v_{j,k}$$
$$= \frac{1}{2\pi |\Sigma_k|^{1/2}} \exp\left(-\frac{1}{2}(X_j - M_k)^T \Sigma_k^{-1} (X_j - M_k)\right) \quad (3)$$

where $X_j = [x_j, B_j^*]^T$, $M_k = [\mu_{x,k}, \mu_{B,k}]^T$ and Σ_k is the covariance matrix of the k^{th} Gaussian cluster in the joint histogram. The means $\mu_{x,k}$ and $\mu_{B,k}$ give the coordinates of the k^{th} Gaussian cluster in the joint histogram. The k^{th} entry of the vector v_j therefore specifies the conditional probability that the j^{th} pixel belongs to the k^{th} cluster, i.e. *given* the current estimate of the pixel intensity x_j, the displacement vector δ_j, and the cluster parameters θ. Note that while we have chosen a Gaussian probability density function to illustrate the method, other bivariate probability density functions could be specified as well.

Next, we impose a spatial coherence constraint on the image x using a Markov Random Field (MRF) distribution. This is achieved using the Gibbsian penalty function

$$P(x) = \frac{1}{Z_x} \exp\left[-\sum_{j=1}^{N} \sum_{n \in \mathcal{N}_j} w_{jn} \phi(x_j - x_n)\right] \quad (4)$$

where Z_x is a normalization constant, n indexes the pixels within each neighborhood $\mathcal{N}_j$ centered on the j^{th} pixel, the convex function $\phi(t)$ penalizes the difference between adjacent pixels of x, and the weights w_{jn} represent the clique weights. In this study we have used an eight nearest neighbor model, and found that a Huber function [4] for $\phi(t)$ gave desirable results. In our initial experiments, we assumed a uniform distrition for $P(x)$ to evaluate the performance of the regularizing term $P(B|x, p, \theta, \delta)$ in isolation, and found that such regularization is already highly effective. However, we also found that adding the Huber prior on x increases the robustness and accelerates the convergence rate of the algorithm (see section III).

The prior probability distribution $P(p)$ is made subject to a similar spatial coherence constraint. More specifically, we vectorize the Gibbsian penalty function as follows

$$P(p) = \frac{1}{Z_p} \exp\left[-\sum_{j=1}^{N} \sum_{n \in \mathcal{N}_j} \left(w_{jn} \sum_{k=1}^{K} \phi(p_{j,k} - p_{n,k})\right)\right] \quad (5)$$

where the meaning of the variables is analogous to that in Eqn. 4. The prior probability distribution $P(\theta)$, in turn, is assumed to be uniform. In practice, a reliable initial estimate of the cluster parameters θ is available in many transmission tomography applications. Hence either fixed values or strong priors could be used to constrain the values of θ. In our experiments we assumed initial guesses for θ that were accurate to within 15%, in which case the algorithm proved to be stable. We plan to further evaluate the algorithm's robustness to the initial guess of θ in future work.

Finally, to achieve spatial smoothness for the deformation field δ, we use a MRF prior defined as

$$P(\delta) = \frac{1}{Z_\delta} \exp\left[-\sum_{p} \sum_{n \in \mathcal{N}_c} w_{cn} \sum_{d=1}^{D} \phi(\delta_{c,d} - \delta_{n,d})\right] \quad (6)$$

where $\mathcal{N}_c$ represents the neighborhood centered on the c^{th} control point (see section II-B.4), and D is the number of spatial dimensions of the deformation map. In this preliminary study, we present results on a 2D phantom. All other variables are analogous to those used in Eqn. 4, with the exception that we here used a quadratic function for $\phi(t)$.

B. Optimization

The posterior probability given in Eqn. 1 can be optimized by maximizing its logarithm, yielding the objective function

$$\psi(x,p,\theta,\delta) = \sum_{i=1}^{M} h_i(l_i) + \beta \sum_{j=1}^{N} \log\left(\sum_{k=1}^{K} v_{j,k} p_{j,k}\right) - \gamma \sum_{C} V_C(x) - \omega \sum_{C} V_C(p) - \eta \sum_{C} V_C(\delta) \qquad (7)$$

where constant terms have been dropped and hyper-parameters have been added to control the strength of each term. Note also that the optimization problem is subject to the nonnegativity constraint for x, and the simplex constraint for p: $p_j \in S_K = \left\{p_j \in \mathbb{R}^K : \sum_{k=1}^{K} p_{j,k} = 1, p_{j,k} \geq 0, k = 1, ..., K\right\}$. In this study, we update the variables x, p, θ, δ sequentially, each time holding the other variables constant. The proposed algorithm cycles through each of these four variables until convergence is reached.

1) x-update: Here the cost function is given by $\psi(x) = \sum_{i=1}^{M} h_i(l_i) + \beta \sum_{j=1}^{N} \log\left(\sum_{k=1}^{K} v_{j,k} p_{j,k}\right) - \gamma \sum_C V_C(x)$. The data match term can be minorized by parabolas of an optimal curvature [3]. Similarly, the second term can be minorized by a sum of quadratics $q_j(x)$, where each $q_j(x)$ has the maximum curvature provided by any of the quadratics $\log(v_{j,k} p_{j,k})$, where $k = 1, ..., K$.[1] This satisfies the conditions for monotonicity, but also creates an unwanted vulnerability to local optima. The problem can be reduced by at every pixel j approximating the Gaussian mixture model $\sum_{k=1}^{K} v_{j,k} p_{j,k}$ by a single Gaussian of the same first and second moments [15]. This approximation becomes progressively more accurate as the weighted sum of Gaussian becomes dominated by one of the component Gaussians. The third term, which consists of a sum of concave functions, was minorized using quadratic surrogates as well. Summed together with the surrogates of the first and second terms, this yields a fully quadratic surrogate to the cost function in x. In the final implementation, the exact and approximate surrogates of the second term were alternated, each time rejecting the approximate surrogate update if it yielded a decrease of the cost function. As confirmed in our experiments (not reported here), this yielded an update scheme that was both monotonic and robust to local optima. The update steps were performed using a steepest ascent scheme.

Note that defining the deformation field δ on x, rather than B, makes the reconstruction of x easier since the data likelihood term is defined in terms of the pixel centers x_j, and not values in between. If the correspondence map were to link the given pixel values B_j to values of x positioned in between the regular pixel grid, we would need to consider details of the image interpolation model when computing the updates for x_j.

[1]Note that this is a conservative choice for the curvature of the surrogates. Computing an optimal curvature is however analytically intractable due to the form of the Gaussian mixture model $\sum_{k=1}^{K} v_{j,k} p_{j,k}$.

2) p-update: In this step, the cost function reduces to $\psi(p) = \beta \sum_{j=1}^{N} \log\left(\sum_{k=1}^{K} v_{j,k} p_{j,k}\right) - \omega \sum_C V_C(p)$. Note that this cost function is entirely concave in p, and bounded only by the simplex constraint on each vector p_j. No surrogates are used for the optimization of p. The direction of steepest ascent within the simplex plane is determined, and the vector field p parameterized as $p = p^q + \lambda p'$, where q is the number of the current iteration, and p' the search direction at that iteration. Taking the first derivative of the resulting function in λ yields a quadratic equation. Hence setting the derivative to zero yields two solutions for λ. One gives a complex value for the cost function in p; the other a real value. Of course, we are each time interested in the latter. The resulting update step is guaranteed to give a monotonic increase in the cost function.

3) θ-update: Here we maximize the function $\psi(\theta) = \sum_{j=1}^{N} \log(v_j \cdot p_j)$ over θ. We used the Nelder-Mead Simplex Method, available as `fminsearch` in Matlab.

4) δ-update: Keeping x, p, and θ constant, the cost function reduces to $\psi(\delta) = \beta \sum_{j=1}^{N} \log\left(\sum_{k=1}^{K} v_{j,k} p_{j,k}\right) - \eta \sum_C V_C(\delta)$, where the first term is the similarity metric, and the second the smoothness penalty. As indicated above, the number of degrees of freedom was reduced by defining a coarse sub-grid of control points on the image x. As this work is in its preliminary stages, we used an off-the-shelf MRF-registration algorithm to update the position of the control points, and hence δ. More specifically, we used an implementation made available by Glocker [2], which uses the fastPD algorithm to optimize MRF-formulated registration metrics using a B-spline deformation model[2]. Glocker's implementation provides a range of similarity metrics, though not the similarity metric derived in Eqn. 2. As a result, we resorted to a standard similarity metric (in this case normalized cross correlation) to perform the δ. This still allowed us to explore the potential of our simultaneous reconstruction and registration approach. Our main objective is to assess the effectiveness of the multiple cluster approximation in regularizing the image reconstruction, and less so in serving as a registration similarity metric. Hence we believe that the use of an alternative similarity metric for the registration step is an acceptable first step towards the development of our SRR algorithm. In our preliminary investigation, the current algorithm converged in all experiments (see section III). We intend to investigate the performance of the proposed similarity metric in the registration component of the algorithm as part of future work.

III. RESULTS

In Fig. 1 we compare the performance of SRR to that of two standard statistical algorithms, namely the unregularized maximum likelihood (ML) algorithm and the penalized ML algorithm with a Huber image prior. Illustrative reconstructions are shown using 16 noiseless projections distributed evenly over an angular range of $\pm 30^\circ$ from the vertical direction. All reconstructions were initialized with a zero image

[2]http://www.mrf-registration.net/.

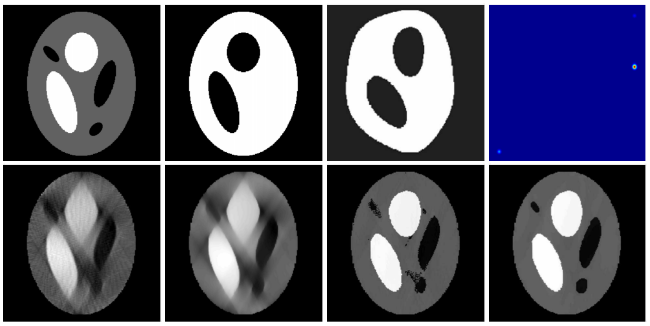

Fig. 1. Top row: phantom, registered prior, nonregistered prior, ground truth joint pdf of phantom and registered prior. The joint pdf contains three clusters (bottom left, center right, top right). Bottom row: unregularized ML, ML with Huber prior on x, SRR without Huber prior on x, SRR with Huber prior on x.

for x, and, where applicable, with a uniform distribution for p, an initial guess within 15% of the ground truth values for θ, and a zero deformation field δ. The phantom and registered as well as non-registered prior are shown in the top row of Fig. 1. Of course, the ground truth registration of the prior is not known to the algorithm. Note that the prior does not contain all of the regions present in the phantom, as would be the case in many real applications. Our method captures such regions in a natural way by modeling them as separate clusters in the joint histogram. The image size was 200x200 for all examples. Compared to the standard ML algorithms, which do not make use of the anatomical prior, the SRR algorithm shows a greatly improved contrast for all regions, even those that were not present in the prior. Note that while SRR already improves the quality of reconstruction when using a uniform prior distribution $P(x)$ (bottom row, third column), its robustness is increased by using a Huber prior for $P(x)$ (bottom row, fourth column). We found that this also increases the convergence rate of the algorithm. Similar results were obtained on a number of different piecewise constant 2D phantoms. The x- and p updates consisted of 50 and 10 steepest ascent steps, respectively, corresponding to approximately 11 and 2 seconds, respectively. The θ updates using the Nelder-Mead Simplex Method required approximately 8 seconds each. Finally, the registration step consisted of 5 iterations of Glocker's fastPD algorithm, which completed in approximately 1 second. Hence the run time per total iteration of the SRR algorithm was approximately 22 seconds. No further attempt was made to accelerate the code. The algorithm converged in approximately 3 update cycles.

IV. CONCLUSIONS AND FUTURE WORK

We proposed a simultaneous reconstruction and registration algorithm that is capable of mitigating the data insufficiency problem of limited view tomography. We derived a cost function using Bayesian probability theory, and proposed a similarity metric based on the explicit modeling of the joint histogram as a sum of bivariate clusters. This proposal followed our previous finding that standard information theoretic similarity metrics are ill-suited for regularizing limited view tomographic reconstructions. So far, we have evaluated the potential of the proposed similarity metric only for regularizing the image reconstruction step, and used an off-the-shelf algorithm with a standard similarity metric (NCC) for the registration step. It is our intention to test the performance of our similarity metric in the registration step at our earliest opportunity. Adherence to a single cost function as given in Eqn. 7 would make a formal characterization of the convergence behavior of the algorithm easier. For example, convergence to either a local or global optimum could be guaranteed by virtue of the boundedness of the cost function when only monotonic updates of the cost function are allowed. Further, we plan to investigate the robustness of the proposed SRR algorithm to measurement noise, over- or underdeclaration of the number of clusters K, registration errors, and sensitivity to variations in the cost function's hyperparameters. Finally, we note that the proposed framework could also be used in the anatomical regularization of hybrid emission tomography applications such as PET/MRI and PET/CT, where the registration component could be used to undo the effects of breathing or other patient motion.

V. ACKNOWLEDGMENTS

Dominique Van de Sompel is supported by a Scatcherd European Scholarship and an EPSRC grant. Michael Brady acknowledges support from DIUS, EPSRC and Siemens.

REFERENCES

[1] B. Nett, J. Tang, S. Leng, and G. Chen. Tomosynthesis via total variation minimization reconstruction and prior image constrained compressed sensing (piccs) on a c-arm system. SPIE, 2008.

[2] B. Glocker, N. Komodakis, G. Tziritas, N. Navab, and N. Paragios. Dense image registration through MRFs and efficient linear programming. *Medical Image Analysis*, 12(6):731–741, 2008.

[3] H. Erdogan and J.A. Fessler. Accelerated monotonic algorithms for transmission tomography. In *ICIP (2)*, pages 680–684, 1998.

[4] P.J. Huber. *Robust Statistics*. Wiley, New York, 1981.

[5] D.J. Godfrey J.T. Dobbins. Digital x-ray tomosynthesis: current state of the art and clinical potential. *Phys. Med. Biol.*, 45:65–106, 2003.

[6] Kak, A. C. and Slaney, M. *Principles of Computerized Tomographic Imaging*. Society of Industrial and Applied Mathematics, 2001.

[7] Y. Mameuda and H. Kudo. New anatomical-prior-based image reconstruction method for pet/spect. *NSS Conf Record*, 2007.

[8] Marroquin, J.L., Santana, E.A., and Botello, S. Hidden markov measure field models for image segmentation. *IEEE trans. on Patt. Anal. and Mach. Intell.*, 25(11):1380–1387, November 2003.

[9] J. Nuyts. The use of mutual information and joint entropy for anatomical priors in emission tomography. *NSS Conf Record*, 2007.

[10] S. Somayajula, E. Asma, and R.M. Leahy. Pet image reconstruction using anatomical information through mutual information based priors. *Nucl. Sci. Symp. Conference Record*, 5:2722–2726, Oct. 2005.

[11] Jing Tang, B.M.W. Tsui, and A. Rahmim. Bayesian pet image reconstruction incorporating anato-functional joint entropy. *ISBI 2008.*, pages 1043–1046, May 2008.

[12] D. Van de Sompel and M. Brady. Robust incorporation of anatomical priors into limited view transmission tomography using multiple cluster modelling of the joint histogram. In *ISBI 2009. Boston, USA.*

[13] D. Van de Sompel and M. Brady. Robust Joint Entropy Regularization of Limited View Transmission Tomography using Gaussian Approximations to the Joint Histogram. In *IPMI 2009. Williamsburg, USA.*

[14] D. Van de Sompel and M. Brady. Simultaneous reconstruction and registration algorithm for limited view transmission tomography using a single bivariate Gaussian approximation to the joint histogram. In *MIUA 2009. London, UK.*

[15] D. Van de Sompel and M. Brady. Simultaneous reconstruction and segmentation algorithm for limited view transmission tomography. *IEEE Transactions on Medical Imaging, 2008. Under revision.*