STATISTICAL FUSION AND SAMPLING OF SCIENTIFIC IMAGES

Azadeh Mohebi and Paul Fieguth

Department of Systems Design Engineering
University of Waterloo, Waterloo, Canada

ABSTRACT

Porous media are an important class of heterogeneous materials possessing complex random structures. Due to the limitation in measuring high resolution real samples, studying different physical properties of porous media requires the reconstruction of artificial samples. In many cases of significant interest, we have a two-scale reconstruction, in which only the large scales are resolved by low resolution measurements, leaving fine scales to be inferred from statistical models. In this paper we propose a statistical fusion approach for a two-scale porous media reconstruction, in which low resolution measurements are fused with high resolution samples, with synthetic realizations generated by posterior sampling.

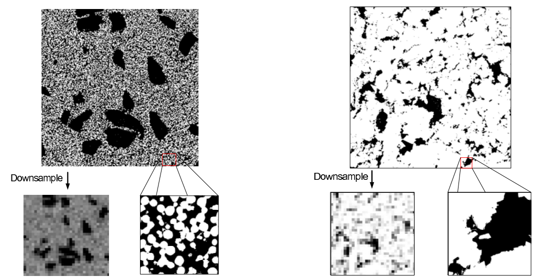
Index Terms— Image sampling, Posterior sampling, Simulated annealing

1. INTRODUCTION

A practical and important class of scientific images are the 2D/3D images obtained from porous materials such as concretes, wood, active carbon, and glass. These materials constitute an important class of heterogeneous media possessing complicated microstructure [1]. Fig. 1 shows two typical 2D samples of porous media: sintered glass spheres and Berea.

Most porous media possess a chaotic structure that is difficult to describe qualitatively [1], but they are not totally random. Therefore there is a mixture of organization and randomness that makes them difficult to characterize and study [1]. In order to study macroscopic properties of permeability, conductivity, and transport, high resolution samples are required. But obtaining such high resolution samples usually requires cutting, polishing and exposure to air, all of which affect the properties of the sample. Moreover, 3D samples obtained by Magnetic Resonance Imaging (MRI) are very low resolution and noisy, so that only the very largest pores are resolved. Therefore, artificial samples of porous media are required to be generated [1].

The reconstruction process is a sampling process. There is a substantial literature, describing this process based on only a prior model, leading to prior sampling [1], [2], [3]. However there is a growing interest in cases in which low-resolution measurements are available, in which case we have a prior—measurement fusion problem, leading to *posterior* sampling. The interest stems from the advancement in 3D magnetic-resonance and 3D computed tomography imaging, where the resulting measurements resolve certain large pores, but leave finer scales unresolved. The fundamental question we are exploring is the degree to which high-resolution 2D scans of similar samples can allow the resolution enhancement of low-resolution 2D or 3D measurement sets, as illustrated in Fig. 2.



(a) Sintered Glass Spheres (b) Berea Carbonate Rock

Fig. 1. Examples of high-resolution 2D porous media (pores are black, solid is white).

2. STATISTICAL FUSION AND MODELING

Our proposed fusion method is based on a Bayesian framework, such that we wish to fuse low-resolution measurements with a high-resolution prior model. Despite a common appearance with super-resolution image reconstruction, we do not have multiple measurements from the same original data and, more importantly, any available high resolution samples do not come from the same sample as the measurements, rather a statistically-equivalent one.

Simple image models, such as correlation models and spatial variance, can not characterize the chaotic and complex morphology of porous media. Instead, for discrete-state problems (porous media images are binary) widely-used Gibbs Random Fields (GRFs) are considered [4]. The Gibbs probability distribution function is defined as

$$p(Z) = \frac{e^{-H(Z)/T}}{\mathcal{Z}} \tag{1}$$

where $H(Z)$ is the *energy function* (prior model) capturing

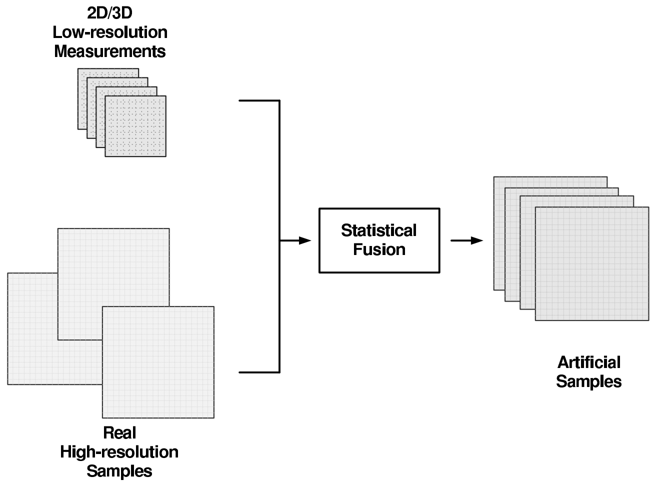


Fig. 2. Statistical fusion of low and high resolution data for porous media reconstruction.

the interaction between neighboring particles, and $\mathcal{Z}$ is a normalization factor.

Given measurements M, the posterior probability distribution is then defined as

$$p(Z|M) = \frac{e^{-H(Z|M)/T}}{\mathcal{Z}_M} \tag{2}$$

$$H(Z|M) = H(Z) + \alpha G(Z; M) \tag{3}$$

where G is just the constraint asserting the measurement in the probability distribution and α is a term balancing the contributions of the prior and measurements.

2.1. Prior Model

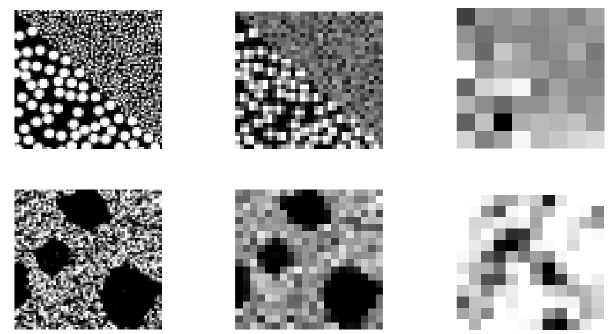
Two common choices for the energy function in porous media literature are chordlength and multipoint (histogram) distribution functions. The chordlength [1] $C^i(\ell)$ is defined to be the probability of finding a chord with length ℓ in phase i. Since porous media contain two phases (pore/black and solid/white) the chordlength distribution can be defined for one or both phases and for chords at different orientations. We have considered dual chordlength model with horizontal and vertical chords. The prior energy for chordlength distribution function is

$$H_c(Z) = \|\bar{C} - C(Z)\| \tag{4}$$

where $\bar{C}$ is the learned / estimated model and $C(Z)$ is the chordlength distribution associated with simulated field Z.

A histogram distribution [3] is non-parametric, keeping the entire joint probability distribution of a local set of pixels within a neighborhood. Choosing eight adjacent pixels as the neighborhood structure leads to a non-parametric model containing a histogram of $2^9 = 512$ probabilities. The prior energy when we consider histogram distribution is

$$H_r(Z) = \|\bar{R} - R(Z)\| \tag{5}$$



(a)Both scales resolved (b)Single scale resolved (c)No scales resolved

Fig. 3. Porosity measurements at different resolutions. Particularly in (c) nearly all information of the image structure has been lost, and the distinction of the two phases in the image is not inferrable from the measurements.

where $\bar{R}$ is the learned / estimated histogram distribution and $R(Z)$ is the histogram distribution of simulated Z.

2.2. Measurements as Constraints

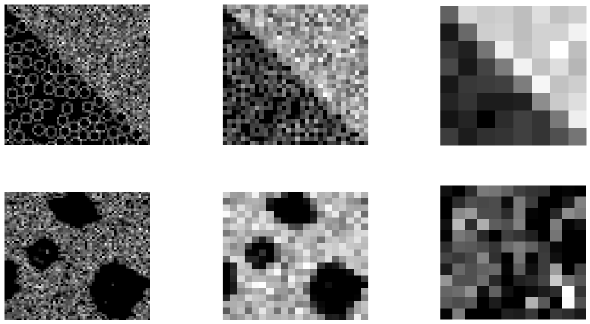
The most common low resolution measurement of a porous medium is the local *porosity*, a measure of the pore space in a material, measured as a fraction between 0 and 1. A low resolution *porosity* measurement (M_p) basically measures porosity of a given material over the dimensions of a 2D pixel or 3D voxel (Fig. 3).

As a complementary measurement, MRI diffusion measurements (M_d) capture information on average pore size, sensitive even to infinitesimal unresolved pores. The measurement is proportional to *Surface-to-Volume* (S/V) ratio. Although low resolution diffusion measurements may also fail to provide detailed information on pore shape, it can provide valuable clues with respect to unresolved structures and geometry, as illustrated in Fig. 4. Parameter d is related to the measurement resolution, such that for $n \times n$ original image and $k \times k$ measurements, $d = \frac{n}{k}$. Apart from the real samples of porous media, we have also shown an artificial example composed of small and large circle to study this type of measurement on a typical two-scale porous media reconstruction problem.

Including both measurement in the Gibbs prior distribution leads to posterior energy

$$\begin{aligned} G(Z; M) &= G(Z; M_d, M_p) \\ &= \|f_d(Z) - M_d\| + \gamma \|f_p(Z) - M_p\| \end{aligned} \tag{6}$$

where $f_d(\cdot)$ and $f_p(\cdot)$ are the forward models for diffusion and porosity measurements, respectively, and γ is a weighting parameter between the two.



(a) $d = 5$ (b) $d = 15$ (c) $d = 25$

Fig. 4. Diffusion measurements at three resolutions. The diffusion measurement encompasses information on the size of different structures at different scales. In sharp contrast to Fig.3, the delineation of the two regions in the top row is clear at even very low resolutions.

3. POSTERIOR SAMPLING

Sampling from the Gibbs probability distribution is straightforward, in principle, by running the Gibbs sampler [5]. We run the Gibbs sampler by annealing, starting with a high temperature (T) which is slowly decreasing. However, in the case of prior sampling the annealing is applied to the single energy function H. The challenge in posterior sampling is the annealing subject to two simultaneous constraints of prior H and measurement function G. In the case of precise measurements, the posterior sampling problem resembles that of annealing subject to hard constraints [5], where the posterior sample would be randomly selected from constrained space

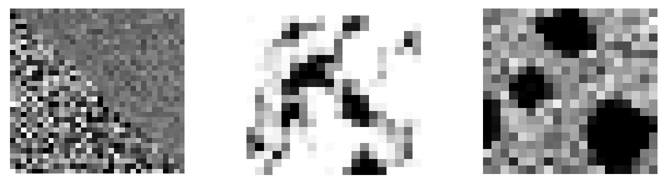
$$\{Z \mid G(Z; M_d, M_p) = 0\}. \tag{7}$$

In the more usual event of unprecise measurements the setting of the relative weight α as a function of annealing iteration is less clear and remains as an open problem.

4. RESULTS AND EVALUATION

We have applied posterior sampling on three types of data, the small-large circle example and two classes of porous materials. In small-large circle example and carbonate rock porous media (Fig. 1, part (b)), chordlength distribution is considered as the prior model, while in sintered glass spheres porous media Fig. 1, part (a)) histogram distribution acts as the prior model. We set the parameters (α and γ) in the posterior energy such that the process starts with the same weight for both measurements and the prior model. To observe the impact of diffusion measurement clearly in the reconstruction process, two types of experiments are done for each type of data: one only with porosity measurement in G, and the other one with both measurements. The results are shown in Fig. 5. We have also shown the reconstructed small scale structures of three different reconstructed samples in Fig. 6.

(a) Original media



(b) Low resolution porosity measurement

(c) Reconstruction, using porosity measurement only

(d) Reconstruction, using both measurements

Fig. 5. Reconstructed artificial samples using constrained sampling, for a degree of downsampling of $d = 15$. Observe the stunning reconstruction, comparing (d) against (b), in the two right columns. The failure to properly reconstruct the left column can be attributed to an inadequate prior.

We have evaluated the results in terms of two aspect: how much the artificial samples are consistent with the measurements and original data, and whether the results contain valid porous media structures.

The correlation between the original data, from which the measurement is generated, and the reconstructed sample is studied as a function of scale. A pixel in the middle of a large pore or solid is likely to be the same in the original and reconstructed images, as opposed to a pixel on a pore/solid boundary. That is, the original-reconstructed correlation is likely a function of structure size. If we measure structure size at a pixel as the number of times a pixel value is unchanged

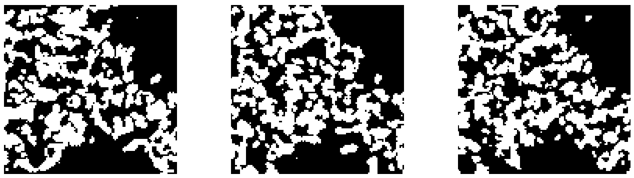
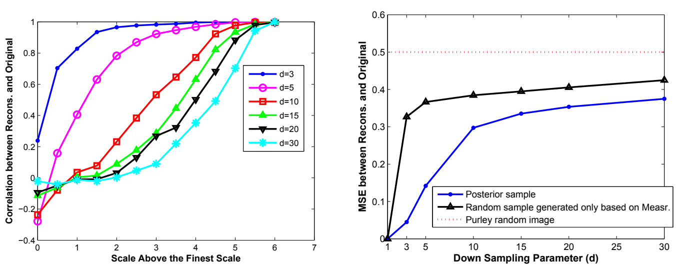


Fig. 6. Three posterior samples: observe that large, resolved structures (such as the pore, right) remain unchanged, whereas unresolved details (fine-scale structure) is randomly synthesized. Thus we construct multiple samples, all representative of a given measured medium.

by sampling, then we can compute correlation as a function of size. Fig. 7 part (a) shows such correlation.

The Mean Squared Error (MSE) between the artificial samples and the original data has been also studied in terms of the resolution of the measurements. Fig. 7 part (b) shows how MSE changes as a function of measurement resolution (d). We can see that after a sharp increase at $d = 10$, MSE does not change significantly as d increases.



(a) Correlation of the recons. with the original as a function of scale

(b) MSE between recons. and original as a function of measurement resolution

Fig. 7. In (a) we plot the correlation of the ground-truth original versus the reconstruction as a function of measurement downsampling and as a function of scale. For several scales below the measured resolution, the posterior reconstruction is positively correlated with truth, meaning that portions of the reconstructed details may be believed. In (b) we plot the MSE between the reconstruction and the original as a function of downsampling d. The three lines show MSE corresponding to a purely random field (top), a prior-free reconstruction based on measurements only (middle), and the posterior sample which we produce (bottom).

To study the similarity and consistency between the artificial and real porous media samples, we have evaluated the artificial reconstructed samples under statistical models learned from real samples. The statistical models represent valid porous media structures and features. The numbers in Table 1 actually shows the dissimilarity between the reconstructed and real samples of porous media in terms of statistically learned models such as chordlength and histogram. According to this table, the artificial samples are more consistent with real porous media when both types of measurements are used as the constraint.

Table 1. Dissimilarity between the artificial and real samples in terms of statistically learned models.

Measurement / **Prior model**	M_d & M_p	M_d	M_p
Chordlength	**0.020080**	0.13050	0.02600
Histogram	**0.000023**	0.00181	0.00016

5. CONCLUSION

In this paper we proposed a statistical fusion approach based on posterior sampling for two-scale porous media reconstruction. In this approach the statistical model learned from the high resolution data is fused with the measurements to construct a posterior model. We have considered two types of measurements. Since the low resolution porosity measurement can only resolve large scale structures, we proposed to add diffusion measurement in the model as well. Diffusion measurement can provide information on the size of structures at different scales. The proposed statistical fusion approach can provide samples which are more correlated and consistent with the real porous media samples.

6. REFERENCES

[1] S. Torquato, *Random Heterogeneous Materials: Microstructure and Macroscopic Properties*, Springer-Verlag, 2002.

[2] M.A. Ioannidis and I. Chatzis, "On the geometry and topology of 3d stochastic porous media," *Journal of Colloid Interface Science*, vol. 229, pp. 323–334, 2000.

[3] S. Alexander and P. Fieguth, "Hierarchical annealing for random image synthesis," *Lecture Notes in Computer Science (EMMCVPR 2003)*, , no. 2683, 2003.

[4] S. B. Gelfand and S. K. Mitter, *Markov Random Fields, Theory and Applications*, Academic Press, 1993.

[5] D. Geman, *Random fields and inverse problmes in imaging*, vol. 1427 of *Lecture Notes in Mathematics*, Springer-Verlag, 1990.

[6] S. Geman and D. Geman, "Stochastic relaxation, gibbs distribution, and the bayesian restoration of images," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, 1984.