Measuring the Distance Between Images and Image Uncertainty Using Wavelet Decompositions and the Earth Mover's Distance

Yunfan Tang[1] and Roy E. Welsch[2, 3]
[1]The University of Hong Kong, Hong Kong, China
[2]Massachusetts Institute of Technology, Cambridge, MA, USA,
[3] Corresponding Author: Roy Welsch, email: rwelsch@mit.edu

Abstract

Most users of statistics are aware of the fact that a different sample from the same population can lead to different sample statistics and often use confidence intervals, margins of error, etc. to show this. However, many users of statistics accept plots and images at face value and pay little attention to showing non-statisticians how they (or the implicit visual models associated with them) might vary from sample to sample. In this paper we discuss ways to measure the distance between images using wavelet decompositions and the Earth Mover's Distance along with various ways to find five image summaries (like five number summaries) to allow quick comparison of central and extreme images. We also consider various ways to model image uncertainty by examining how one might bootstrap image residuals after the wavelet structure has been removed. Since there is always some structure left in the residuals, we face many of the same problems (but with more complexity) that are faced when bootstrapping time series residuals after building a time series model (sieve bootstrap). We demonstrate our techniques on a series of fMRI brain images.

Keywords: bootstrap, fMRI, image decomposition and uncertainty, visualization,

1. Introduction

It is now quite common to consider images as a type of data (Big Data) that might be amenable to statistical analysis and modeling. Recent advances in technology have dramatically lowered the cost of image production and thus generated many image databases from surveillance cameras, medical imaging equipment, etc. Under such circumstances, it is often of great interest to see whether statistical methodologies for numerical samples can carry over to image samples. In particular, we are interested in seeking an efficient approach to represent and visualize sampling variability for images. Bootstrapping methods for numerical samples have been well studied in the past. Researchers began to extend the traditional bootstrapping techniques to samples with spatial pattern since Hall (1985) and Buhlmann and Kunsch (1995) first explored block bootstrapping. Lahiri (2003) provided a comprehensive review of spatial bootstrapping techniques. Whitcher (2006) proposed a resampling method for images through wavelet packet decomposition. He used a spatial autocorrelation test to find sub-images (coefficients) with spatial randomness in the wavelet packet decomposition and then bootstrapped on these sub-images. Menjoge (2010) and Menjoge and Welsch (2010) suggested bootstrapping on scatterplots and used the Earth Mover's Distance as a metric to compare the bootstrapped samples. This allowed visualizing the sampling variability of scatterplots through a few summary plots such as the central plot and the most distant pair of plots.

Most statistical analyses of images treat them as spatial data on a two-dimensional lattice. In the pixel representation, an image with an $N \times N$ array of pixels would have a very high dimensionality of N^2. In addition, the complexity of image objects leads to heavy spatial correlation among pixels, which causes dependence problems in statistical algorithms such as regression or bootstrapping. Therefore, it is often desirable to use sparse representations for images to deal with the huge computational cost and dependency structure. For example, principal component analysis (PCA) and singular value decompositions (SVD) are commonly used in image compression, facial recognition, texture analysis, etc. A detailed description of these methods can be found in Diamantaras and Kung (1996). There are also a number of time-frequency representation methods in the signal processing literature such as wavelet and Fourier transforms (Mallat 2008).

The main goal of this paper is to visualize the sampling variability of images through wavelet bootstrapping and the Earth Mover's Distance. Section 2.1 and 2.2 provide a short background on the Earth Mover's Distance and wavelet packet transform, respectively. Section 3 explains algorithms for wavelet bootstrapping and sets up a pairwise metric for bootstrapped samples, followed by various ways to use image statistics for visualization. Section 4 demonstrates the application to a brain MRI image and discusses empirical parameter choices for computational issues. Section 5 concludes the paper and suggests some ideas for future work.

2. Related work

2.1 Earth Mover's Distance

The Earth Mover's Distance (EMD) was proposed by Werman et al. (1985) as a metric between histograms or distributions. It is also called the Wasserstein metric in mathematics and the Mallows distance in statistics. Intuitively, EMD regards each bin of a histogram as a pile of dirt and measures the total amount of work to transform one histogram into another. EMD could be calculated through the transportation problem as the uncapacitated minimal cost flow with computational cost $O(N^3 log(N))$ (Ahuja et al. 1993). In computer vision, EMD is first used as a metric between gray-scale images by Peleg et al. (1989). In their case, EMD is computed on the image domain, where bin values on the two-dimensional histogram are determined by the intensity of pixels. Rubner et al. (2000) gave a formal definition of EMD for histograms and signatures and extended it as an image metric to color images for content-based retrieval. They also discussed that EMD matches perceptual dissimilarity better than L_1 and χ^2 distance.

2.2 Wavelet packet analysis

In the discrete wavelet transform, the original signal is first passed through a low-pass filter to obtain approximation coefficients and through a high-pass filter to obtain detailed coefficients. On the next level, it is only the approximation coefficients that are again decomposed by these filters (Mallat 2008). In contrast, wavelet packet analysis decomposes both the detailed coefficients and approximation coefficients on each level. For a detailed description of the discrete wavelet transform and wavelet packet analysis, see Mallat (2008). Wavelet packet analysis allows more flexibility in partitioning the discrete signal, since there are no limitations on the wavelet bases. This enables us to find the best representations as an alternative to the standard wavelet transform. Because of these advantages, it has been applied to time series bootstrapping by Percival et al. (2000) and Whitcher (2001) and to images by Whitcher (2006).

3. Algorithm

For image bootstrapping, we mainly follow the idea of Whitcher (2006), but adopt slight modifications. In his top-down approach, sub-images at each level of a wavelet packet decomposition are tested for spatial autocorrelation by Moran's I (Moran 1950) or Geary's C (Geary 1954). If one sub-image is declared to have spatial randomness, it will be sent to bootstrapping and not decomposed further. In our case, we fully decompose the entire image into a wavelet packet tree at a pre-fixed level and obtain a set of sub-images $\{W_{i,j},\ 1 \le i, j \le 2^n\}$, n being the level of decomposition. Spatial autocorrelation tests are conducted on all $W_{i,j}$ and sub-images with spatial randomness are resampled through naïve bootstrapping methods. In other words, the bootstrap procedure is only applied to wavelet coefficients with no inner-correlation, ensuring the preservation of the basic spatial structure by the rest of the coefficients with dependency structure. Finally, we reconstruct the whole image using the inverse wavelet transform of all sub-images.

To quantify the pairwise distance between bootstrapped images, we again apply wavelet packet decompositions to both images X^1 and X^2 and obtain $\{W^1_{i,j},\ 1 \le i, j \le 2^n\}$ and $\{W^2_{i,j},\ 1 \le i, j \le 2^n\}$. Then we use EMD to measure the distance between every pair of wavelet sub-images $(W^1_{i,j}, W^2_{i,j})$ with the same index following the definition of Rubner et al. (2000). Notice that they only define EMD on non-negative histograms, but wavelet coefficients on sub-images can assume negative values. One solution is to take their absolute values before comparison, but this implies that the images to be compared are altered as well. Our method is to shift all the coefficients in both images to the non-negative domain by adding a minimal constant. This intuitively comes from the fact that EMD between histograms with equal total sum is invariant to shifting. After calculating the total flow cost between two sub-images, we divide the result by min(sum($|W^1_{i,j}|$), sum($|W^2_{i,j}|$)) to avoid favoring smaller images, where the sum is taken over the absolute value of coefficients on each sub-image before shifting.

After setting up the pairwise metric, we use the idea from Menjoge (2010) to obtain summary statistics for visualization of the sampling variability. These include a central image as the one with the smallest total sum of distance to all other images, the two images with largest distance from each other, and the ranked order of all images using a travelling salesman algorithm (Lawler et at. 1985). These could be generalized as follows:

1. Given an image sample X, compute all sub-images $\{W_{i,j},\ 1 \le i, j \le 2^n\}$ through a wavelet packet transform of level n.
2. For each $W_{i,j}$, conduct spatial randomness tests using Moran's I or Geary's C.
3. Bootstrap on the $W_{i,j}$ with spatial randomness and obtain $W^*_{i,j}$.
4. Conduct an inverse wavelet transform to obtain X^*.
5. Repeat step 1 - 4 many times.
6. Calculate pairwise distances between bootstrapped samples. For images X^1 and X^2, first apply the wavelet packet decomposition to obtain $\{W^1_{i,j},\ 1 \le i, j \le 2^n\}$ and $\{W^2_{i,j},\ 1 \le i, j \le 2^n\}$. Then $d(X^1, X^2) = \Sigma_{i,j}\,\mathrm{EMD}(W^1_{i,j},\ W^2_{i,j})$ on the shifted coefficients.
7. Find the image summary statistics according to the pairwise metric.

4. Example

The brain MRI image we choose is subject Brown_26001, taken from ADHD-200 database of 1000 Functional Connectomes Project by NITRC. We crop and resize the original image from 240 × 256 to 138 × 138 and conduct four-level wavelet packet analysis using Daubechies wavelet 4 "db4" in MATLAB. A total of nine images are bootstrapped using the algorithms described above. In step 3, we need to set up a threshold for Moran's I. Sub-images with the absolute value of Moran's I lower than the threshold are declared to have spatial randomness and sent for bootstrapping. The optimal threshold could be found through trial and error for the most appealing visual effect. Empirically, bootstrapped samples will become overly altered and unrecognizable if the threshold is set too large since the fundamental image structure is preserved within certain wavelet coefficients that should not be changed. In our experiment, 49 of 256 sub-images with lowest absolute value of Moran's I are bootstrapped.

Computational complexities of our algorithms mainly arise out of the wavelet packet transform and EMD calculation. Obtaining a level L fully decomposed wavelet packet tree of an image with size $N \times N$ has complexity $O(LN^2\log(N))$ in the worst case scenario, which could be conducted very efficiently by most modern CPUs. The EMD algorithm takes another $O(n^3\log(n))$ where n is the total number of pixels in the sub-image. In MATLAB the level-4 sub-image decomposed by the "db4" wavelet has size 15×15 so n equals 225, and all of the 49 selected bootstrapped sub-images need to be compared. The average time to calculate each pairwise metric is 4.02 seconds on our machine with an Intel i7-3610 2.3 GHz CPU and 8GB RAM.

Figure 1 shows the original image and two bootstrapped images with longest distance from each other. Our algorithm generates bootstrapped samples with varying textures, brightness and contrasts in different parts of the brain. Most significant spatial alternations in these two images occur in the parietal lobe and occipital lobe (the mid-left and mid-upper section of the image) where the abundance of sulcus indicates potential variations. In Figure 2 we rank the nine bootstrapped images in order by the travelling salesman algorithm so as to minimize the total sum of adjacent images, producing a gradual shift in visual perception from one extreme to another. These statistics altogether could provide helpful predictions of possible variations of the original image.

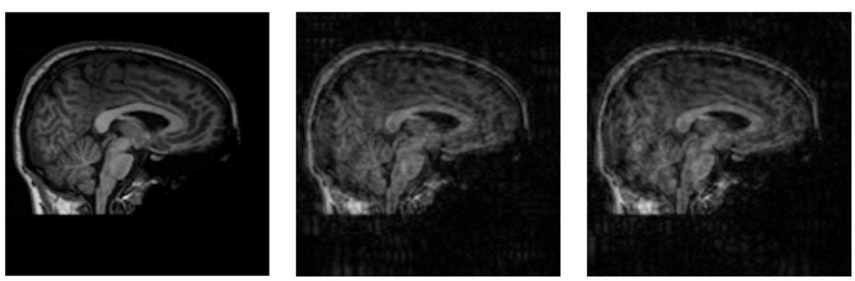

Figure 1. Original Image (left) and two bootstrapped images farthest from each other.

5. Conclusions and Future Work

In this paper we have addressed ways to order a collection of images and, by so doing, prepare summaries that might include the extreme and middle images, for example. One way such a collection might be generated is by thinking of an image as being a sample from a population of images. If we just have a single image and want to understand something about how sampling uncertainty affects that image, then the bootstrap is one possible approach. However, bootstrapping data with structure, such as a time series or image is difficult. One idea is to remove most of the structure and then bootstrap what is left. We have done this by using wavelet decompositions of the image to obtain residuals with minimal structure. The bootstrapped residuals are then added back to the wavelet decomposition to obtain a new image and this process is repeated many times by using repeated bootstrapping of the residuals to obtain a collection of images that can then be ordered and summarized in some way.

The Earth Mover's Distance has been used to measure distance between images. In the future we plan to explore other possible distance measures perhaps more related to the content and applications intended for the images. We have also assumed the images in a collection of images are independent of each other. In many situations this is not true and we plan to consider how to summarize correlated images.

6. Acknowledgments

This work was supported in part by the Singapore-MIT Alliance for Computation and Systems Biology and the MIT Center for Computational Research in Economics and Management Science. We also would like to thank Rajiv Menjoge for making some of his programs available for our use.

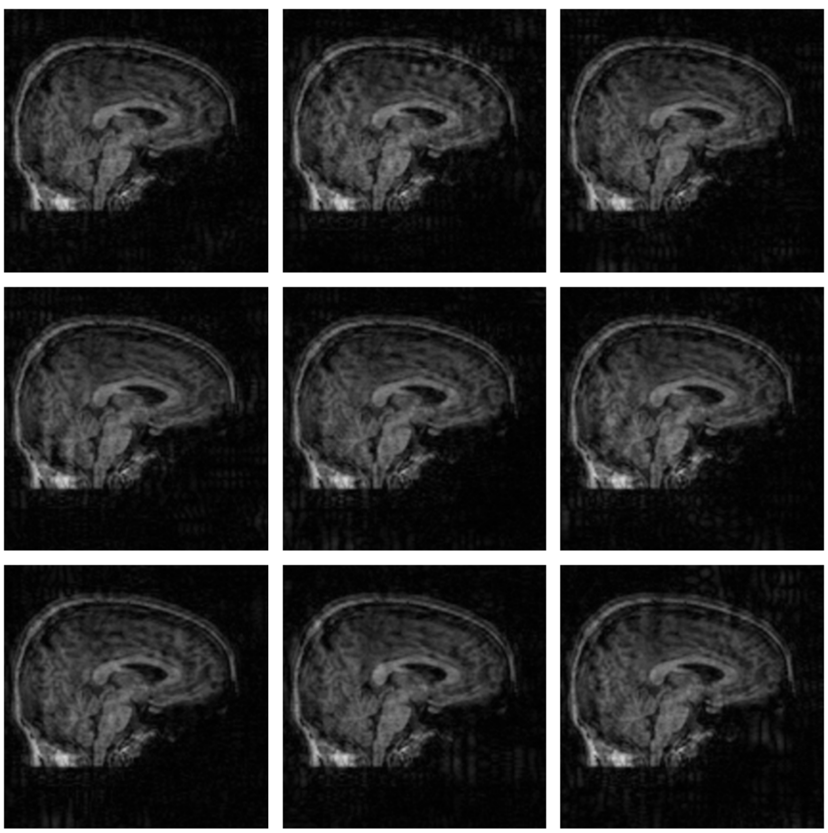

Figure 2. Bootstrapped images ranked by the travelling salesman problem from top-left to bottom-right.

References

Ahuja, K., Magnanti, T.L., and Orlin, J.B (1993) *Network Flows*, Prentice Hall: Englewood Cliffs.

Buhlmann, P. and Kunsch, H.R. (1995) "The blockwise bootstrap for general parameters of a stationary time series," *Scandinavian Journal of Statistics*, 22, 35-54.

Diamantaras, K.I. and Kung, S.Y. (1996) *Principal Component Neural Networks: Theory and Applications,* John Wiley & Sons.

Geary, R. (1954) "The contiguity ratio and statistical mapping," *The Incorporated Statistician*, 5, 115-145.

Hall, P. (1985) "Resampling a coverage pattern," *Stochastic Process,* 20(2), 231-246.

Lahiri, S.N. (2003) *Resampling Methods for Dependent Data*, Springer Series in Statistics.

Lawler, E.L., Lenstra, J.K., Kan, A.H.G. Rinnooy and Shmoys, D.B. (1985) *The Traveling Salesman Problem*, John Wiley & Sons.

Mallat, S. (2008) *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way,* Academic Press.

Menjoge, R. (2010) "New procedures for visualizing data and diagnosing regression models," Ph.D. Thesis, Operations Research Center, MIT Sloan School of Management.

Menjoge, R. and Welsch, R.E. (2010) "Visualizing the Sampling Variability of Plots," *Proceedings of COMPSTAT 2010*, 1375-1382, Springer.

Moran, P.A.P. (1950) "Notes on continuous stochastic phenomena," *Biometrika* 37, 17-23.

Peleg, S., Werman, M., and Rom, H. (1989) "A unified approach to the change of resolution: space and gray-level," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11, 739–742.

Percival, D.B., Sardy, S., and Davision, A.C. (2000) "Wavestrapping time series: adaptive wavelet-based bootstrapping," *Nonlinear and Nonstationary Signal Processing*, Cambridge University Press, 442-471.

Rubner, Y., Tomasi, C. and Guibas, L.J. (2000) "The Earth Mover's Distance as a metric for image retrieval," *International Journal of Computer Vision,* 40(2), 99-121.

Werman, M., Peleg, S., and Rosenfeld, A. (1985) "A distance metric for multidimensional histograms," *Computer Vision, Graphics and Image Processing*, 32(3), 328–336.

Whitcher, B. (2001) "Simulating Gaussian stationary processes with unbounded spectra," *Journal of Computational and Graphical Statistics*, 10(1), 112-134.

Whitcher, B. (2006) "Wavelet-based bootstrapping of spatial patterns on a finite lattice," *Computational Statistics and Data Analysis*, 50(9), 2399-2421.