

Classifications of Patterns for Multiple Point Statistical Simulation

Hai T. Nguyen and Jeff B. Boisvert

Multiple-point statistical (MPS) simulation uses statistics directly from a training image (TI) to generate conditional realizations that contain the complex features in the TI. MPS is usually implemented to generate categorical variable models, the comparison between patterns found in the TI and patterns in the realization is straightforward. Extending MPS simulation to continuous variables requires a calculation of the difference between two patterns consisting of continuous values at the data locations and the TI. Considering the squared difference or the correlation between patterns is likely ideal; however, for the number of comparisons made in a typical simulation, this can be CPU demanding for large models. Five different methods for comparing patterns are explored.

Introduction

Training images (TI) are used to represent the statistical properties of a region. MPS algorithms aim to impose the multivariate statistics of the TI on the desired conditioning data in a geostatistical realization. In MPS, the first step is to group training patterns together using a difference measure. T. Zhang (2006) presented a 6 or 9 filter score algorithm to classify a pattern in 2 D and 3 D for MPS simulation. It is similar to pattern matching in image processing.

Several difference measures are explored in this work to compare patterns in the TI with local conditioning data found during MPS simulation. The following measures are considered (1) mean difference value (2) sum of square difference (SSD) (3) 9 and 15 point filters score and (4) correlation. The typical difference measure used is the SSD or the sum of the absolute difference; however, both measures require significant CPU time for large patterns. Alternative measures will be considered if they rank patterns in a similar way as the SSD method, but require fewer computations. Different TIs with various pattern sizes in 2D and 3D are tested. The interested reader is referred to paper 105 in this report for the discussion of an MPS simulation algorithm for continuous variables.

Pattern Difference Metrics

Five pattern difference metrics are discussed for measuring the difference between patterns.

Mean Difference: Based on the difference between the mean value of a pattern in MPS simulation and the TI pattern, the difference value (Δ) can be defined as: $\Delta = |m_{pat} - m_{TI}|$; where: m_{pat} and m_{TI} are mean value of the real pattern and TI pattern respectively. It is unlikely that the mean will be a good measure of the difference between patterns but it was included as a base case for the other difference metrics.

SSD Rank: The equation for SSD matching algorithm is:

$$SSD = \sum_{i=1}^n (P_i^{Pat} - P_i^{TI})^2 \quad (1)$$

where: SSD is the sum of square difference, n is the number points (pixels) of each pattern, P_i^{Pat} P_i^{TI} is value of cell i of the real pattern and TI pattern, respectively. The SSD is likely the ideal metric as it considers the difference in all values in the pattern. If the TI is sufficiently large so that there exists a pattern with similar values to P_i^{Pat} this metric should locate the correct pattern.

Filters score based: the filter scores based approach for training pattern classification was introduced by Zhang et al. (2006). The approach works with both categorical and continuous TI. A filter is defined over a pattern which can be presented as a set of weights attached to each node. Each score is considered as a summary of a pattern. For illustration, Figure 1 shows the process of calculating a filter score for a given 2D filter (J. B. Wu). Each filter represents different spatial aspects of the pattern.

In 3D, the r^{th} score $S_r(i, j, k)$ of each training pattern center at location $u=(i,j,k)$ in the TI and the related filter $f_r(x,y,z)$ is:

$$S_r(i, j, k) = \sum_{z=-m}^m \sum_{y=-m}^m \sum_{x=-m}^m f_r(x, y, z) T(i+x, j+y, k+z) \quad k=1, \dots, 9 \quad (2)$$

Where: $S_r(i, j, k)$ the filter score value; (i, j, k) is the coordinate of the center node u in the data template; x and y , and z vary from $-m$ to m ; $2m+1$ is the number of nodes in the X, Y, Z directions respectively. Nine filters are defined in the 3D condition as the following (T. Zhang 2006):

- f_1 filter, North-South average, which highlights the center of a pattern:

$$f_1(x, y, z) = 1 - \frac{|y|}{m} \in [0, 1], y = -m, \dots, +m \quad (4)$$

- f_2 filter, East-West average, which can be obtained by rotating the f_1 filter by 90 degrees:

$$f_2(x, y, z) = 1 - \frac{|x|}{m} \in [0, 1], x = -m, \dots, +m \quad (5)$$

- f_3 filter, North-South gradient, which detects the edge of a pattern

$$f_3(x, y, z) = \frac{y}{m} \in [-1, 1], y = -m, \dots, +m \quad (6)$$

- f_4 filter, East-West gradient, which can be obtained by rotating the f filter by 90 degrees:

$$f_4(x, y, z) = \frac{x}{m} \in [-1, 1], x = -m, \dots, +m \quad (7)$$

- f_5 filter, North-South curvature, which detects the North-South pattern curvature:

$$f_5(x, y, z) = \frac{2|y|}{m} - 1 \in [-1, 1], y = -m, \dots, +m \quad (8)$$

- f_6 filter, East-West curvature, which can be obtained by rotating the f_5 filter by 90 degrees:

$$f_6(x, y, z) = \frac{2|x|}{m} - 1 \in [-1, 1], x = -m, \dots, +m \quad (9)$$

- f_7 filter, Top-Bottom average:

$$f_7(x, y, z) = 1 - \frac{|z|}{m} \in [-1, 1], x = -m, \dots, +m \quad (10)$$

- f_8 filter, Top-Bottom gradient:

$$f_8(x, y, z) = \frac{|z|}{m} \in [-1, 1], x = -m, \dots, +m \quad (11)$$

- f_9 filter, Top-Bottom curvature:

$$f_9(x, y, z) = \frac{2|z|}{m} - 1 \in [-1, 1], z = -m, \dots, +m \quad (12)$$

For a specific case with $m = 1$, calculation of 15 filter scores in the 3D condition from f_1 to f_{15} filter are shown in the Figure 2. In order to measure the similarity between two patterns, the difference between filter scores of the real pattern and TI pattern are defined as:

$$Filter.Dif_9 = \sum_{k=1}^9 (S_{p_{l-k}} - S_{TI-k})^2 \quad (13)$$

15-Filter scores based approach: this ranking method is similar to the 9-filter scores. However, we include 6 additional filters from f_{10} to f_{15} which can be replaced for the gradient and curvature filters in the case of missing data points in a pattern. The equation for calculating the difference between filter scores of two patterns is similar to equation 10:

$$Filter.Dif_{15} = \sum_{k=1}^{15} (S_{pt_k} - S_{TI_k})^2 \tag{14}$$

For 9 and 15 Filter score calculation, we assume that a new pattern will be generated which is based on the original one as showing in the Figure 3. Individual filters are dropped if there are missing data points in the pattern.

Correlation Rank: The correlation coefficient can be applied to measure the similarity of a given pattern and patterns in the TI. The coefficient of correlation between the current real pattern, Pat_i and the TI pattern in the database is denoted T_i which is computed by the following equation (Normal Cross Correlation):

$$\rho_{(Pat,T_i)} = \frac{\sum_{u=1}^{xsz} \sum_{v=1}^{ysiz} \sum_{w=1}^{zsz} [(P_{(u,v,j)} - m_p)(TI_{i+u,j+v,k+w} - m_{TI_i})]}{\sum_{u=1}^{xsz} \sum_{v=1}^{ysiz} \sum_{w=1}^{zsz} [(P_{(u,v,j)} - m_p)^2 \sum_{u=1}^{xsz} \sum_{v=1}^{ysiz} \sum_{w=1}^{zsz} (TI_{i+u,j+v,k+w} - m_{TI_i})^2]} \tag{15}$$

In the equation $\rho_{(Pat,T_i)}$ is the correlation coefficient; n are numbers of rows and columns in the pattern and m_x, m_y are mean of corresponding of current real pattern Pat_i and the TI patterns, T_i. The correlation coefficient will also be computationally expensive to calculate but is included for comparison.

Results

The TI used to compare difference metrics in this study is shown in Figures 4 and 5 with dimensions 128x115x11. Figure 6 shows the different template patterns considered; the goal is to locate similar patterns in the TI. By implementing pattern classification using Mean, SSD, 9 filters, 15 filters and the correlation, the results show that a method for image matching based on SSD algorithm is able to produce accurate results for different patterns. To qualify the others, scatter plot is used to present the correlation between two different methods (Figures 7 to 12), summarized in Tables 1 and 2:

Table 1. Summary of comparison between SSD and other methods using different templates

Measure	Pattern 1		Pattern 2		Pattern 3		Pattern 4		Pattern 5		Pattern 6	
	Corr.	Top 100*	Corr.	Top 100	Corr.	Top 100	Corr.	Top 100	Corr.	Top 100	Corr.	Top 100
Mean	-0.125	1	0.208	8	0.1	3	0.167	2	0.027	7	-	6
9 Filter	0.447	19	0.282	38	0.269	40	0.197	22	0.091	27	-	28
15 Filter	0.212	24	0.261	45	0.021	50	0.184	31	0.063	31	-	28
Correlation	0.542	NA	0.559	NA	0.754	NA	0.562	NA	0.757	NA	0.902	54

Note: * Numbers of TI patterns on top 100 rank of SSD Rank and one of 4 above methods

Table 2. Mean of Correlation coefficient & Average # of TI patterns of six patterns on top 100 ranks between SSD and four others

Measurement methods	For 6 patterns	
	Mean of Corr. Coef.	Average # of TI patterns on top 100 ranks
Mean Dif.	0.239	5
9 Filter	0.563	29
15 Filter	0.619	35
Correlation	0.134	9

Based on the results of the five different methods for a given TI with six different patterns, it can be concluded that the SSD, 9-filter, and 15 filter scores produce similar results for different patterns. Table 1, 2 and Fig. 7 – 12 show that the 15 filter score approach has the best fit with the SSD algorithm. Both the mean difference and correlation are dissimilar to the SSD method. However, when the pattern has a large number of points, the correlation produced similar results (Figure 12 for a pattern size of 9x9x9).

This is expected as the correlation would not be stable for a small number of points in a pattern. The summary of five different classification methods is shown in Table 3.

Table 3. Summary of five different methods, advantages and drawbacks of each one

Measure	Algorithm	Advantages	Disadvantages
Mean Difference	$\Delta = m_{pat} - m_{TI} $	- Low CPU time	-generates many patterns with similar difference metric -Difference measure is not sensitive to local variations
SSD	$SSD = \sum_{i=1}^n (P_i^{Pat} - P_i^{TI})^2$	-Accurately measure local differences between patterns	-CPU intensive to calculate
10 Filter scores	- Define filters and filter scores - $Filter.Dif_9 = \sum_{k=1}^9 (S_{pt-k} - S_{TI-k})^2$	- Good result for all patterns - Quick due to the reduction of calculations	- Not as sensitive to local changes as SSD
15 Filter scores	- Define filters and filter scores - $Filter.Dif_{15} = \sum_{k=1}^{15} (S_{pat_i} - S_{TI-k})^2$	-Good result for all patterns -Quick due to the reduction of calculations	-Not as sensitive to local changes as SSD
Correlation	$\rho_{(Pat, TI)} = \frac{\sum_{u=1}^{size} \sum_{v=1}^{size} \sum_{w=1}^{size} [(P_{(u,v,j)} - m_p)(TI_{i+u,j+v,k+w} - m_{TI})]}{\sum_{u=1}^{size} \sum_{v=1}^{size} \sum_{w=1}^{size} [(P_{(u,v,j)} - m_p)^2] \sum_{u=1}^{size} \sum_{v=1}^{size} \sum_{w=1}^{size} [(TI_{i+u,j+v,k+w} - m_{TI})^2]}$	- Good result for larger patterns.	- Time consuming to calculate. - Unstable for low numbers of data.

Conclusions

In this paper, five pattern difference metrics are compared for use in MPS simulation. The SSD algorithm, 9-filter score, 15 filter score and correlation perform reasonably well and will be implemented in full MPS simulation and tested for CPU and pattern similarity performance.

References

Du and Zhang, 3D Porosity Simulation of Porous Media Using Continuous Multiple-point *Geostatistics, Intelligent Computing and Intelligent Systems*, 376 – 380 (2009)

Han and Fan. Mesh Pattern Recognition Using Correlation Matching Method, . In *Proceedings of MVA*. pp.383-386 (1994).

Deutsch, C.V. and Journel, A.G., 1998, *GSLIB: Geostatistical Software Library and User’s Guide*, Oxford University Press, New York, 2nd Ed., 369 pp.

Zhang, "Filter-based training pattern classification for spatial pattern simulation," Ph.D. dissertation, USA: Stanford University, 2006, pp. 11-26.

Zhang, Bombarde, Strebelle, and Oatney, "3d porosity modeling of a carbonate reservoir using continuous multiple-point statistics simulation," 2005, SPE paper # 96308, pp. 1-6.

Zhang, Switzer, and Journel, Filter-Based Classification of Training Image Patterns for Spatial Simulation, *Mathematical Geology*, Vol. 38, No. 1, January 2006.

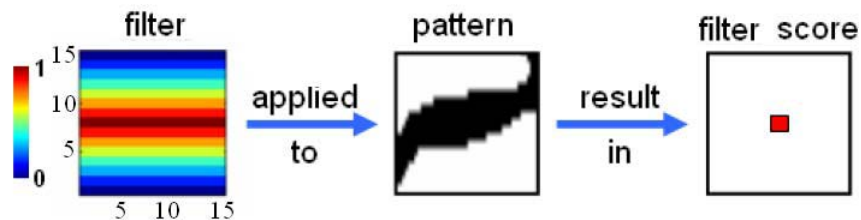


Figure 1. Illustration of obtaining a filter score of a training pattern (Zhang).
 (a) a filter; (b) a training pattern; (c) a filter score.

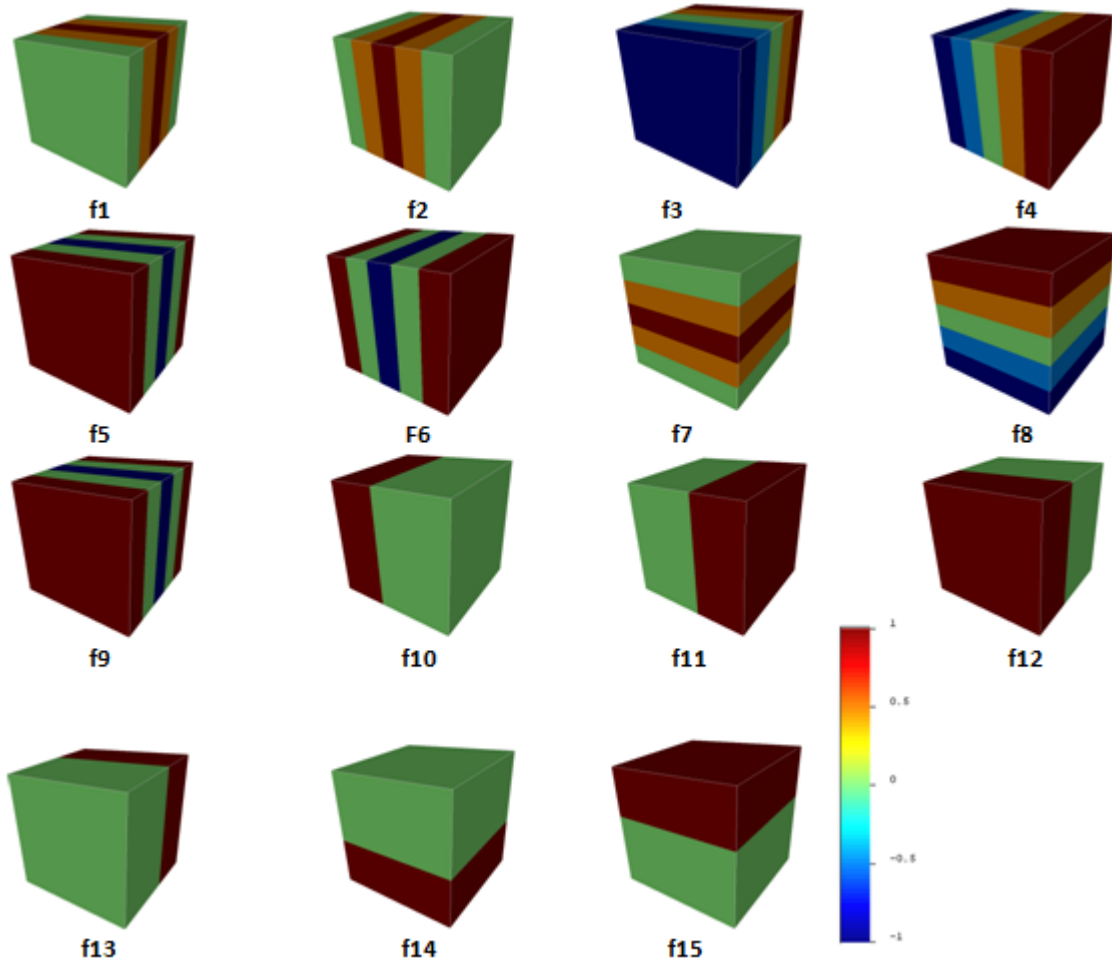


Figure 2: illustration of 15 filters used.

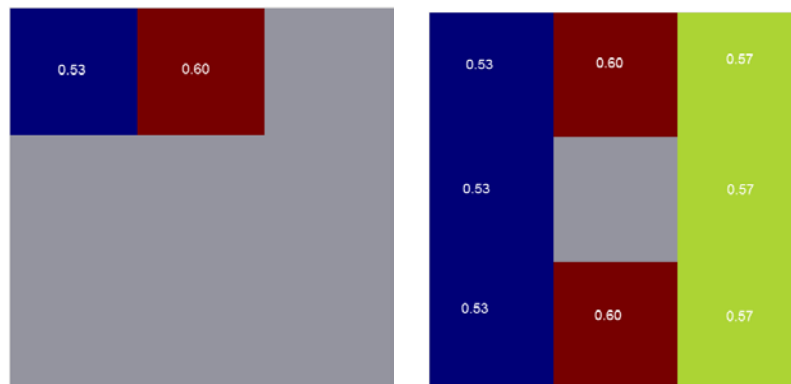


Figure 3. Updating missing data points of the pattern. Left: a pattern in the realization. Right: Values are inferred to generate an exhaustive pattern for comparison to the TI.

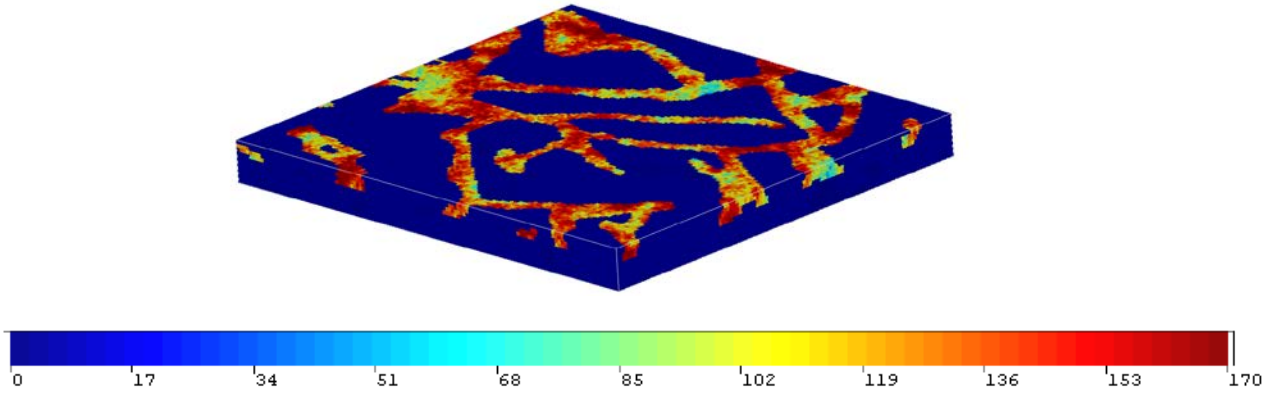


Figure 4: 3 D Model from Training Image ($128 \times 115 \times 11$). Units are not specified but can be considered related to grade.

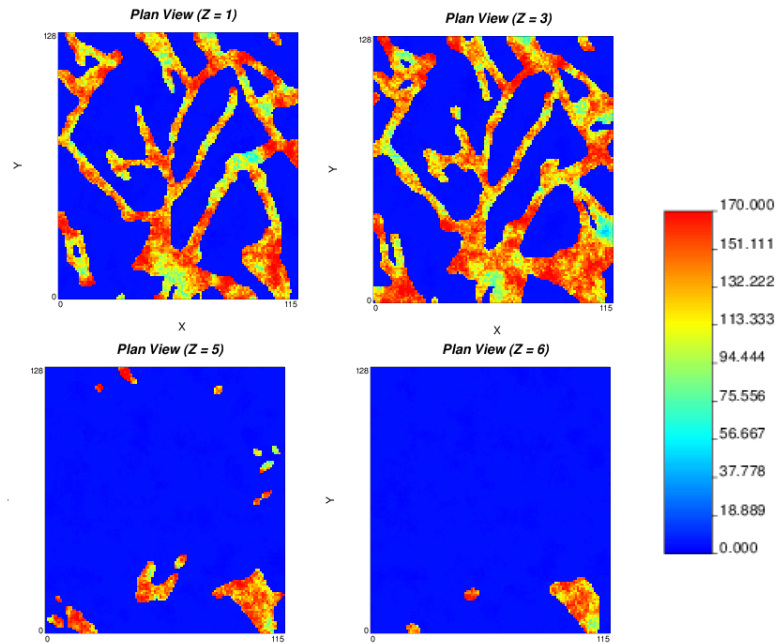
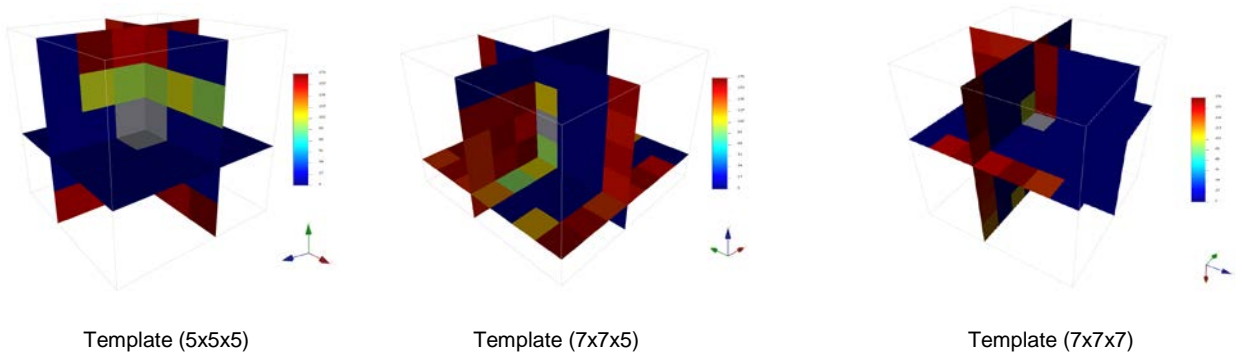


Figure 5. Sections of the TI



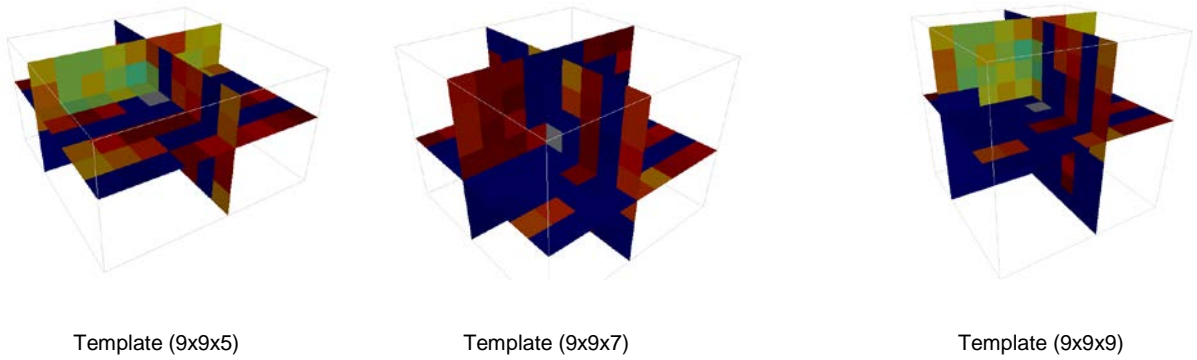


Figure 6: Different Templates

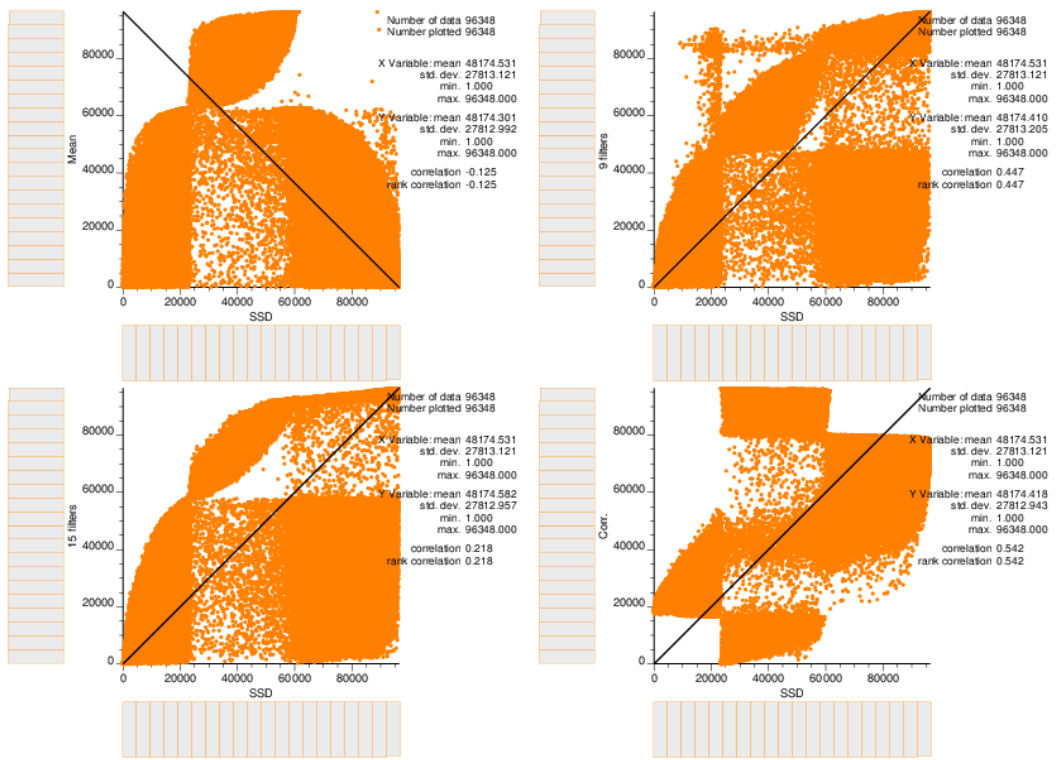


Figure 7: Correlation between SSD Rank and four other methods for the template size of 5x5x5

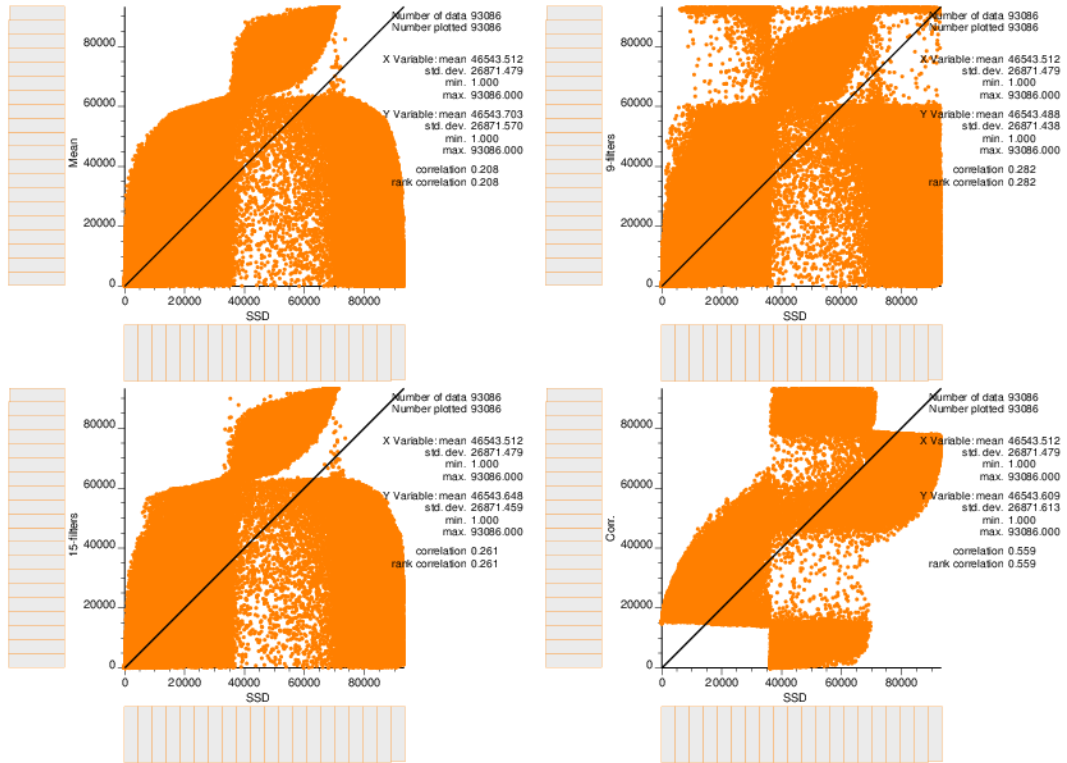


Figure 8: Correlation between SSD algorithm and four other methods for the pattern size of 7x7x5

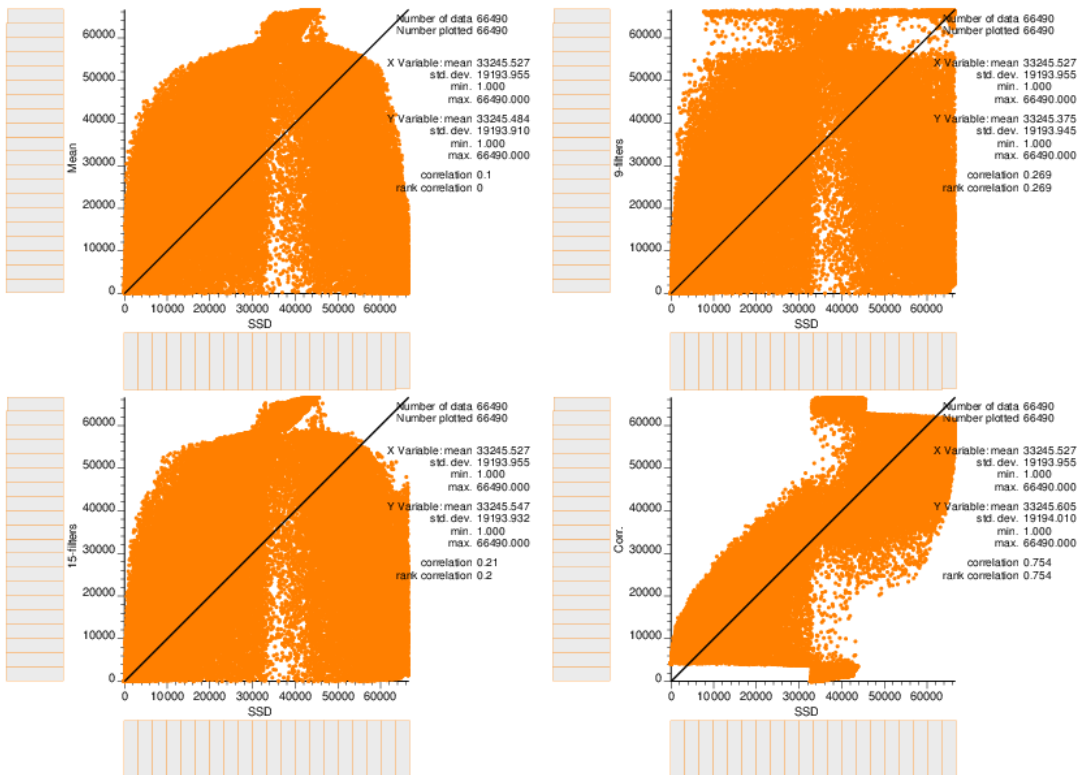


Figure 9: Correlation between SSD algorithm and four other methods for the pattern size of 7x7x7

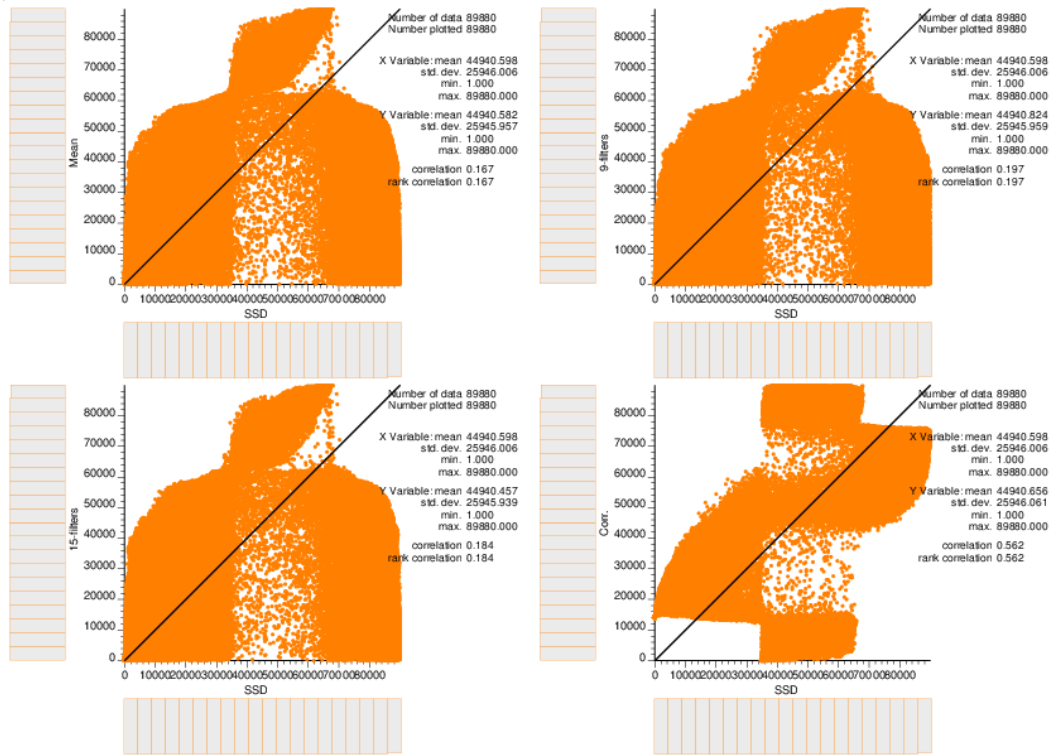


Figure 10: Correlation between SSD algorithm and four other methods for the pattern size of 9x9x5

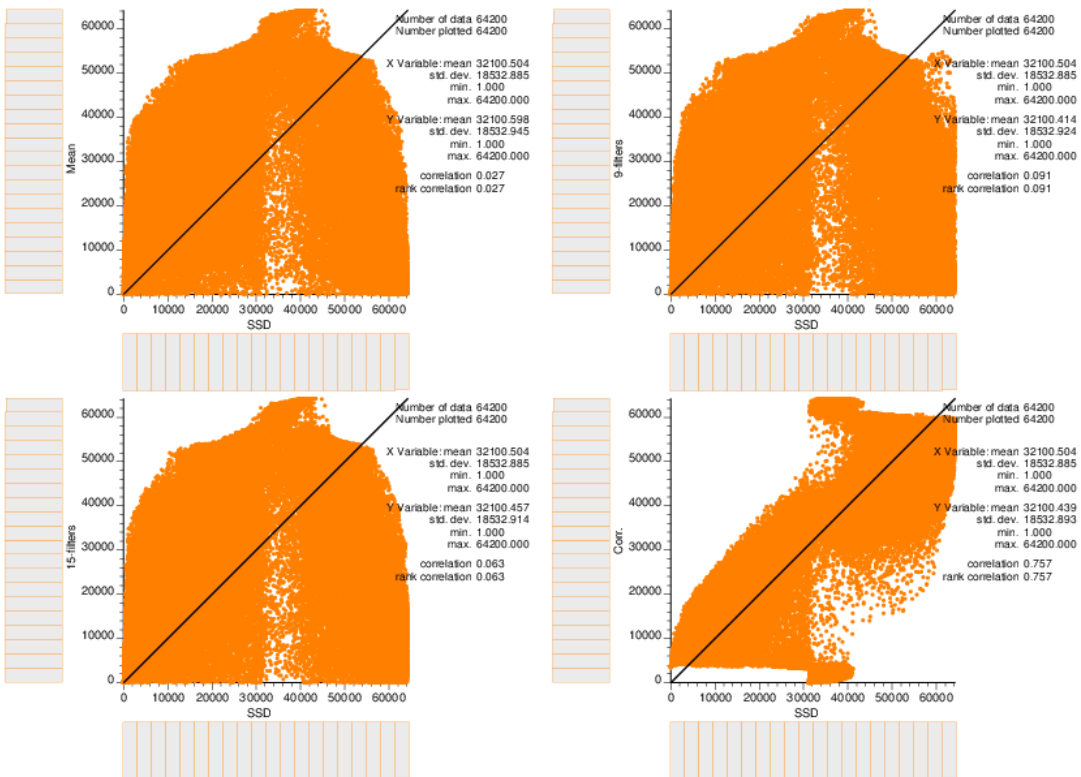


Figure 11: Correlation between SSD algorithm and four other methods for the pattern size of 9x9x7

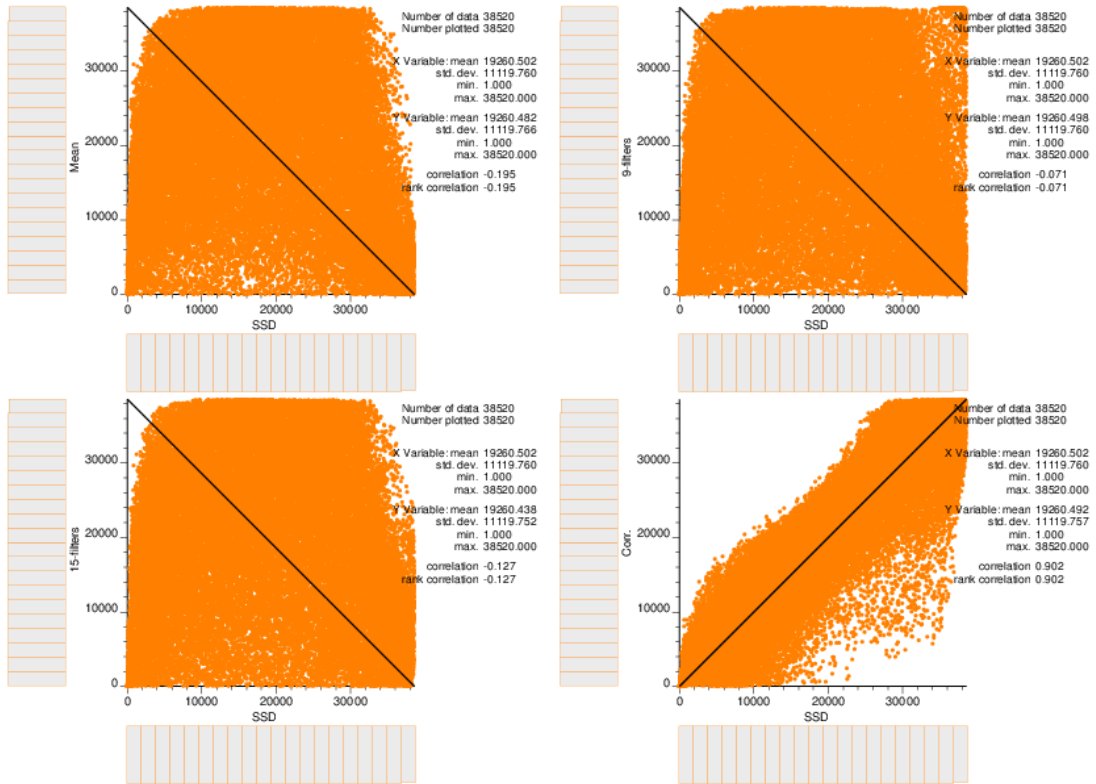


Figure 12: Correlation between SSD algorithm and four other methods for the pattern size of 9x9x9