Chapter 28
Optimum Scale Selection for 3D Point Cloud Classification through Distance Correlation

Manuel Oviedo de la Fuente, Carlos Cabo, Celestino Ordóñez and Javier Roca-Pardiñas

Abstract Multiple scale machine learning algorithms using handcrafted features are among the most efficient methods for 3D point cloud supervised classification and segmentation. Despite their proven good performance, there are still some aspects that are not fully solved, determining optimum scales being one of them. In this work, we analyze the usefulness of functional distance correlation to address this problem. Specifically, we propose to adjust functions to the distance correlation between each of the features, at different scales, and the labels of the points, and select as optimum scales those corresponding to the global maximum of said functions. The method, which to the best of our knowledge has been proposed in this context for the first time, was applied to a benchmark dataset and the results analyzed and compared with those obtained using other methods for scale selection.

28.1 Introduction

In recent decades there has been an explosion of sensors and techniques to obtain spatial data representing real objects by means of 3D point clouds. Laser scanners, either static, mobile, portable or airborne, as well as cameras and computer vision algorithms, especially the Structure-from-Motion (SfM) algorithm [13], are currently the main sources of this kind of data. From the beginning, it was quite evident that there was a need to develop algorithms for the automatic extraction of useful information from the point clouds; a need that increased with the progressive capacity

Manuel Oviedo de la Fuente (✉)
Universidade de Santiago de Compostela, Spain, e-mail: manuel.oviedo@usc.es

Carlos Cabo
Universidad de Oviedo, Spain, e-mail: cabo.gmail@uniovi.es

Celestino Ordóñez
Universidad de Oviedo, Spain, e-mail: ordonezcelestino@uniovi.es

Javier Roca-Pardiñas
Universidade de Vigo, Spain, e-mail: roca@uvigo.es

© Springer Nature Switzerland AG 2020

G. Aneiros et al. (eds.), *Functional and High-Dimensional Statistics and Related Fields*,
Contributions to Statistics, https://doi.org/10.1007/978-3-030-47756-1_28

of the sensors to measure larger point clouds each time. Among these algorithms, those based on the application of machine learning techniques have proven to be very efficient [10, 4], and, accordingly, their use is very extended nowadays. Moreover, their efficiency increases when the features (covariates) of the model are not extracted at a single scale but at several scales. In practice, this means that feature extraction is carried out considering different sizes of the neighborhood (scale) around each point (or voxel, when the point clouds is simplified by means of voxelization). In this way, the extracted features for a point (voxel) at different scales capture different characteristics of the objects around that point, and this helps in the classification procedure. Unfortunately, the selection of the scales is often carried out heuristically, taking into account the density of the point cloud, the kind of objects to be classified, the noise of the data, etc. On other ocassions, the procedure simply consists in selecting a number of scales at regular intervals. These procedures are quite objective, and have some drawbacks [7]. For that reason, it is important to carry out research into more objective scale selection methods, as an adequate selection has a positive influence on the results of the classification. Previous works have addressed this problem from different perspectives. One of them is to find the scale for which the labelling of the current point is the most similar to the labellings of its neighbors at the same scale [4]. Another approach [15] estimates the optimum scales taking into account the local structure of the covariance matrix and the Shanon entropy [11]. In this work, we propose a different approach that assumes that the optimum scales should correspond to the local maximum of the functions obtained calculating the distance correlation between each of the features at a number of scales (i.e. 100) and the values of the labels.

28.2 Methodology

28.2.1 Feature Extraction

A key aspect of machine learning applied to point cloud segmentation and classification is to define and determine the features (input variables) to be introduced in the mathematical models. The multi-scale strategy is based on the fact that a region around a point can look like a 1D, 2D or 3D object depending on the size of the region [2]. The input variables (features) included in the supervised classification algorithms are algebraic expressions involving the eigenvalues of the eigendecomposition of the local covariance matrix $\mathbf{\Sigma}$: $\Sigma = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{p}_i - \bar{\mathbf{p}})^T(\mathbf{p}_i - \bar{\mathbf{p}}) = \mathrm{V}\Lambda\mathrm{V}^{\mathrm{T}}$, where $\mathbf{p}_i = (X_i, Y_i, Z_i)$ is a point of the point cloud, $\lambda_1 > \lambda_2 > \lambda_3$ are the eigenvalues, V a matrix whose columns are the corresponding eigenvectors, and N the number of points inside a sphere of center $\mathbf{p}_i$ and radius R. That is, the eigenvalues, and consequently the features extracted from the point cloud, depend on the values of the scale (radius of the sphere).

The relationship between the values of the eigenvalues $\lambda_1, \lambda_2, \lambda_3$ at a point is related to the local geometry at that point [5]: a linear 1D structure when $\lambda_1 \geq \lambda_2, \lambda_3$; a planar 2D structure when $\lambda_1, \lambda_2 \geq \lambda_3$ and a volumetric 3D structure when

$\lambda_1 \approx \lambda_2 \approx \lambda_3$. Specifically, the features extracted for each point at each scale are: *Linearity* $L = (\lambda_1 - \lambda_2)/\lambda_1$, *Planarity* $P = (\lambda_2 - \lambda_3)/\lambda_1$, *Sphericity* $S = \lambda_3/\lambda_1$, *Horizontality* $H = acos(\mathbf{v}_3 \cdot \mathbf{z})/\|\mathbf{v}_3\|$ and *Z range* $Z = Z_{max} - Z_{min}$, which are very common in the literature [15]. Calculation of Z range for each point is not limited to a sphere but to a vertical cylinder of a specific section (scale) around that point. In order to avoid the negative effect of outliers, Z coordinates are limited to an interval between the 5th and 95th percentiles.

28.2.2 Optimum Scale Estimation from Distance Correlation Functions

Distance correlation [12] is a measure of the degree of correlation, linear or non-linear, between two variables of arbitrary finite dimensions. When the data are ordered and close enough, it is possible to approximate distance correlation values for functions, and analyze them using methods for functional data. Particularly, we are interested in determining the global maximum of the distance correlation function, as it is supposed to be the point that captures the most relevant information concerning the relationship between the variables. A similar approach was used by the authors for variable selection in regression and classification problems [1, 6, 9]. $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$ being two random vectors, distance correlation between X and Y is defined as

$$\mathcal{R}^2(X,Y) = \begin{cases} \frac{\mathcal{V}^2(X,Y)}{\sqrt{\mathcal{V}^2(X)\mathcal{V}^2(Y)}}, & \mathcal{V}^2(X)\mathcal{V}^2(Y) > 0 \\ 0, & \mathcal{V}^2(X)\mathcal{V}^2(Y) = 0 \end{cases} \tag{28.1}$$

where $\mathcal{V}^2(X,Y) = ||f_{X,Y} - f_X f_Y||^2 = \frac{1}{c_p c_q} \int_{\mathbb{R}^{p+q}} \frac{|f_{X,Y}(t,s) - f_X(t) f_Y(s)|^2}{|t|^{1+p}|s|^{1+q}}$ is the distance covariance, a measure of the distance between $f_{X,Y}$, the joint characteristic function of random vectors X and Y, and the product $f_X f_Y$ of the characteristics functions of X and Y, respectively. For their part, c_p and c_q are constants depending on the dimensions p and q, respectively.

Distance correlation has some advantages over other correlation coefficients, such as the Pearson correlation coefficient. First, it measures non-linear dependence. Second, X and Y do not need to be one dimensional variables. Third, $\mathcal{R}(X,Y) = 0 \Leftrightarrow X, Y$ are independent, that is, independence is a necessary and sufficient condition for the nullity of distance correlation.

Once the correlation distance has been determined for each feature at different scales $k \in \mathbb{R}$, we adjust a function $m : k \rightarrow \mathcal{R}(X,Y)$, and determine the values of k corresponding to global maximum of this function. Then, different supervised classification algorithms are applied using the features at those scales, and the results compared with those obtained when features are calculated at a specific number of scales at constant intervals or following an exponential function.

28.3 Experimental Results

28.3.1 Dataset

In order to evaluate the performance of the proposed methodology, we apply it to the Oakland 3D point cloud dataset [8], a benchmark dataset that has been previously used in different studies concerning point cloud segmentation and classification. The 3D point cloud was collected around the CMU campus in Oakland - Pittsburgh (USA) using a Mobile Laser Scanner (MLS), that consists of two-dimensional laser scanners, an Inertial Measurement Unit (IMU), and a Global Navigation Satellite system (GNSS), all of them calibrated and mounted on the Nablab 11 vehicle. Figure 28.1 shows a small part of the point cloud, where six labels have been marked.

Fig. 28.1 Small area of the Oakland point cloud dataset. Each point has been assigned a label.

28.3.2 Neighborhood Selection

Consider a sample data $\{\mathbf{X}_i, \mathbf{G}_i\}_{i=1}^{n}$ where $\mathbf{X} = \left(X^1, ..., X^{J=5}\right)$, represents the vector of features (*linearity*, *planarity*, *sphericity*, *horizontality* and *Z range*), and $\mathbf{G} = (G_1, .., G_{m=5})$ the vector of classes (**cars**, **buildings**, **canopy**, **ground** and **poles**). For each sample i, each feature is evaluated at a regular grid of $N = 100$ scales measured in centimetres: $X_i^j = \left(X_i^j(t_1), X_i^j(t_2), \ldots, X_i^j(t_N)\right)$. Figure 28.2 shows a sample of $n = 150$ curves for each features registered in the interval $k \in [t_1 = 50, t_{100} = 300]$ and the corresponding functional mean, both colored by class label. Note the different performance of the features for the different classes and scales. For instance, *horizontality* takes high values for the **ground**, and it is uniform at different scales. However, this feature shows abrupt jumps at certain scales for the **poles**, that could correspond to edge effects. As expected, *linearity* takes high values for the **poles** and low values for the **buildings**.

Figure 28.3 shows the distance correlation functions for 100 repetitions of random samples of size $n = 750$ (150 per class), corresponding to each of the features extracted. A histogram of the global maximum of distance correlation curves for those repetitions is depicted at the bottom of the figure. As can be appreciated, most of the maxima (impact points) correspond to low scales, except for the *Z range* variable (5th - 95th range of z axis).

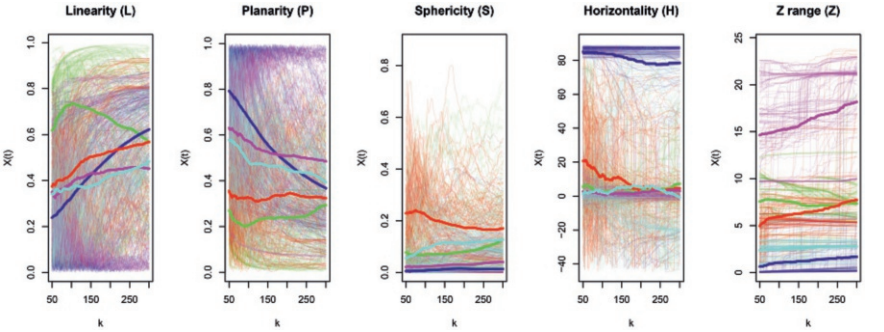

Fig. 28.2 A sample of features curves: poles (green), ground (blue), vegetation (red), buildings (magenta), and vehicles (cyan). Functional means for each class are represented as wider lines.

Our aim is to estimate an optimum neighborhood (scale) for each feature by means of distance correlation (DC), taking into account its advantage with respect to the Pearson coefficient. On the one hand, we calculated the distance correlation between the dependent variable (the label for each curve) and each of the features, see Figure 28.3. On the other hand, we calculated the distance correlation between the labels and two independent variables, *horizontality* and *Z range*, given that these features are more correlated with the dependent variable. In addition, DC between the dependent variable and the five features was also calculated. It is evident that DC functions are not very different for the five features analysed, and it is also evident that maximum values are reached at low scales, except for the range of z. In addition, it can also be appreciated that DC for *horizontality* and *Z range* are significantly

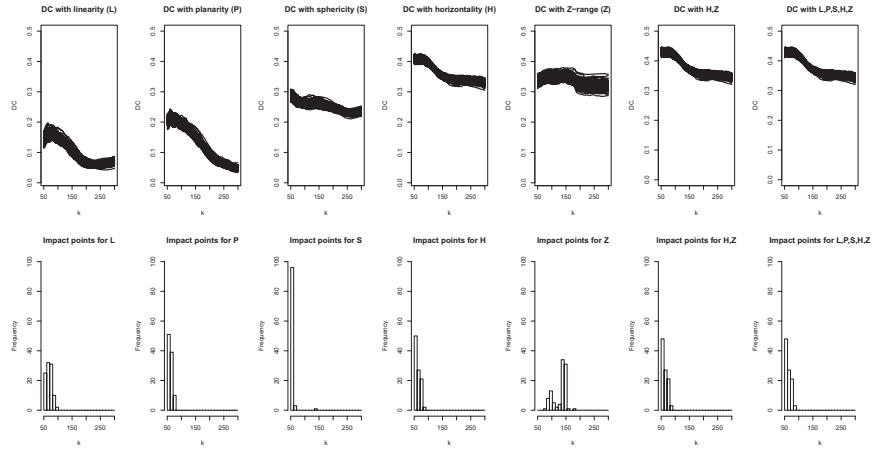

Fig. 28.3 Distance correlation functions between the group class and the features curves (top) and histogram with the scale for the global maximum on each function (bottom).

Table 28.1 Metrics of the classification using logistic regression (LR) and random forest classifier (RF) for different scenarios of k.

		Precision %					Recall %					F1 %				
Model	k	Poles	Grou.	Veg.	Build.	Cars	Poles	Grou.	Veg.	Build.	Cars	Poles	Grou.	Veg.	Build.	Cars
LR	50	70	96	74	**88**	76	74	**99**	72	77	82	72	98	73	82	78
LR	100	63	96	70	74	80	78	98	49	78	80	70	98	57	76	80
LR	175	51	96	66	64	75	63	93	32	78	82	57	95	43	70	78
LR	237	37	97	57	53	60	49	90	30	79	43	42	93	39	63	49
LR	300	28	95	49	49	46	34	88	32	80	25	31	91	39	61	31
LR	k_λ	61	97	70	74	77	73	96	50	78	80	66	97	75	75	52
LR	k_{mdc_2}	72	97	75	87	77	77	99	72	77	82	75	98	73	82	79
LR	k_{mdc_5}	74	97	75	**88**	78	76	**99**	71	79	87	75	98	73	**83**	**82**
RF	50	74	97	76	76	77	74	**99**	73	79	76	74	98	74	77	76
RF	112	**76**	**99**	**78**	71	**85**	**80**	96	74	79	77	**78**	98	76	78	81
RF	175	58	96	66	66	81	67	91	53	73	79	62	94	59	69	80
RF	237	52	96	57	67	76	63	89	45	68	79	57	92	50	67	77
RF	300	51	96	52	58	75	51	86	43	61	**89**	51	90	47	59	81
RF	k_λ	70	98	73	80	83	74	95	68	77	78	72	96	70	73	80
RF	k_{mdc_2}	**76**	98	**78**	74	82	77	98	77	**80**	77	76	98	77	77	78
RF	k_{mdc_5}	**76**	**99**	**78**	75	84	75	**99**	**80**	78	78	75	**99**	**78**	76	81

higher than for the other three features, which suggests that DC might be used not only to estimate the optimum scales but also to select the most important features to be included in the classification models.

In order to contrast the performance of this approach, we followed the proposal of [15] computing the optimal scale k_λ, corresponding to the minimum of the Shannon entropy E_λ, which depends on the normalized eigenvalues $e_i, i = 1, ..., 3$, of the local covariance matrix Σ: $E_\lambda = -e_1 ln(e_1) - e_2 ln(e_2) - e_3 ln(e_3)$ (2).

28.3.3 Classification

The scale corresponding to the most frequent values providing a global maximum (impact points) was used as input variable for two classification algorithms, multinomial logistic regression classifier (LR) and random forest classifier (RF) [14], in two scenarios: (a) k_{mdc_2}: only the features with the highest distance correlation values (*horizontality* and *Z range*) were included in the model and (b) k_{mdc_5}: all the features (*linearity*, *planarity*, *sphericity*, *horizontality* and *Z range*) were used to train the models. Additionally, we used the following values of the scale k: (c) k_λ, obtained according to equation (2), (d) k_{seq}, linearly spaced scales corresponding to the following values of k in centimetres (cm): 50, 112, 175, 237, 300 and (e) k_{exp}, exponential spaced scales, that corresponds to $k = 1, 3, 7, 20, 55, 300$ cm. This last option arises from the fact that the global maximums of DC correspond to low scales.

Training data (150 per class) and test data (500 per class) were sampled from different areas of the point cloud, in order to ensure their independence. Table 28.1 shows the results of the classification for the test sample, in terms of precision, recall and F1-score, for each of the scales, using a logistic regression (LR) and random forest classifier (RF).

The metrics for the classification have quite different values depending on the category, see Table 28.1. Thus, the best results were obtained for **ground** class, followed by **cars**. Lower values were obtained for **poles**, **vegetation** and **buildings**. The results are very similar when the set of the five features (k_{mdc_5}), or just the two with the highest values of distance correlation (*horizontality* and *Z range*), are included in the model (k_{mdc_2}). In general, the models that use the scales corresponding to maximum distance correlation outperform the others, including that corresponding to the minimum of the Shannon entropy (k_λ), that did not turn out to be particularly good. Table 28.2 shows a decrease of the accuracy classification with the scale in both classifiers. So, it is better to calculate the scales using exponential function k_{exp} than using a linear function k_{seq}. This approach limits the number of scales to be calculated, thus reducing computing time.

Table 28.2 Total accuracy in % of the classification using logistic regression (LR) and random forest (RF) classifiers for sequential k_{seq}, exponential k_{exp}, k_λ, k_{mdc_2} and k_{mdc_5} scales.

	k_{seq}	50			100	150		200	250	300			
Model	k_{exp}	50	60	70	100		190			300	k_λ	k_{mdc_2}	k_{mdc_5}
LR		81	81	81	79	72	68	67	56	51	74	81	**83**
RF		80	81	82	81	75	72	71	66	65	78	81	81

28.4 Conclusions

Selecting optimum scales for supervised classification of 3D point clouds is relevant not only to improve the results but also to understand the effect of the features involved in the classification when the local neighborhood changes. We assume as hypothesis of our study that calculating the maximum of the distance correlation functions between the features (input variables) and the classes (output variable) can help to determine the optimum scale for classification and to select the most important variables at that scale. This hypothesis was tested on a benchmark 3D point cloud from an urban environment, and the analysis of the results indicates that our approach outperforms other common methods for scale selection, in particular one that uses specific scales at regular intervals and another that calculates the optimum scale using Shannon's information. Moreover, the analysis of the distance correlation functions for the different features provides information about the importance of these features in the classification. The best results were obtained when the five features, calculated at the optimum scale, were included in the classification model, but similar results were obtained when only the two features with the highest values of the correlation distance were considered. Accordingly, distance correlation function could be used as a filter for feature selection regardless of the classification algorithm. For future work, we plan to analyse a multi-scale analysis using significant structures of the features curves [3, 6].

Acknowledgements M. Oviedo acknowledges support from MTM2016-76969-P project (Spanish Ministry of Science, Innovation and Universities and the European Regional Development Fund).

References

[1] Berrendero, J.R., Cuevas, A., Torrecilla, J.L.: Variable selection in functional data classification: a maxima-hunting proposal. Statistica Sinica, 619–638 (2016)

[2] Brodu, N., Lague, D.: 3D terrestrial lidar data classification of complex natural scenes using a multi-scale dimensionality criterion: Applications in geomorphology. ISPRS Journal of Photogrammetry and Remote Sensing **68**, 121–134 (2012)

[3] Chaudhuri, P., Marron, J.: Scale space view of curve estimation. Ann. Stat, 408–428 (2000)

[4] Demantké, J., Mallet, C., David, N., Vallet, B.: Dimensionality based scale selection in 3d lidar point clouds (2011)

[5] Dittrich, A., Weinmann, M., Hinz, S.: Analytical and numerical investigations on the accuracy and robustness of geometric features extracted from 3D point cloud data. ISPRS Journal of Photogrammetry and Remote Sensing **126**, 195–208 (2017)

[6] Febrero-Bande, M., González-Manteiga, W., Oviedo de la Fuente, M.: Variable selection in functional additive regression models. Computational Statistics **34**(2), 469–487 (2019)

[7] Mallet, C., Bretar, F., Roux, M., Soergel, W., Heipke, Ch.: Relevance assessment of full-waveform lidar data for urban area classification. ISPRS Journal of Photogrammetry and Remote Sensing **66**(6), 571–584 (2011)

[8] Munoz, D., Bagnell, J., Vandapel, N., Hebert, M.: Contextual classification with functional max-margin Markov networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 975–982 (2009)

[9] Ordóñez, C., Oviedo de la Fuente, M., Roca-Pardiñas, J., Rodríguez-Pérez, J.L.: Determining optimum wavelengths for leaf water content estimation from reflectance: A distance correlation approach. Chemometrics and Intelligent Laboratory Systems **173**(15), 41–50 (2018)

[10] Pauly, M., Keiser, R., Gross, M.: Multi-scale feature extraction on pointsampled surfaces. Comput. Graph. Forum **22**(3), 281–289 (2003)

[11] Shannon, C.: A mathematical theory of communication. Bell Syst. Tech. J. **27**, 379–423 (1948)

[12] Székely G.J., Rizzo, M.L., Bakirov, N.K.: Measuring and testing dependence by correlation of distances. Annals of Statistics **35**(6), 2769–2794 (2007)

[13] Ullman, S.: The Interpretation of Structure from Motion: Proceedings of the Royal Society of London. Series B, Biological Sciences **203**(1153), 405–426 (1979)

[14] Venables, W.N., Ripley, B.D.: Modern Applied Statistics with S. Springer (2002)

[15] Weinmann, M., Jutzi, B., Hinz, S., Mallet, C.: Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. ISPRS Journal of Photogrammetry and Remote Sensing **105**, 286–304 (2015)