A STEGANALYTIC ALGORITHM FOR 3D POLYGONAL MESHES

Ying Yang[⋆] *Ruggero Pintus*[⋆,†] *Holly Rushmeier*[⋆] *Ioannis Ivrissimtzis*[‡]

[⋆] Department of Computer Science, Yale University [†] CRS4, Italy
[‡] School of Engineering and Computing Sciences, Durham University

ABSTRACT

We propose a steganalytic algorithm for watermarks embedded by Cho et al.'s mean-based algorithm [1]. The main observation is that while in a clean model the means of Cho et al.'s normalized histogram bins are expected to follow a Gaussian distribution, in a marked model their distribution will be bimodal. The proposed algorithm estimates the number of bins through an exhaustive search and then the presence of a watermark is decided by a tailor made normality test. We also propose a modification of Cho et al.'s algorithm which is more resistant to the steganalytic attack and offers an improved robustness/capacity trade-off.

***Index Terms*—** Polygonal meshes, watermarking, data embedding, steganalysis

1. INTRODUCTION

A *digital watermark* is a digital signal embedded into a digital medium to protect it from unauthorized use or alteration. Most of the existing methods for embedding invisible watermarks on 3D models are mostly concerned with the robustness of the watermark, targeting applications such as proof of ownership and copy control. However, we believe that, alongside robustness, undetectability should also be a major concern when measuring the performance of such algorithms.
Contribution: In Section 2, we propose a steganalytic algorithm for detecting the presence of a watermark hidden by Cho et al.'s mean-based algorithm [1]. To the best of our knowledge, it is the first specific 3D steganalytic method proposed in the literature and our tests show that it outperforms the universal algorithm in [2]. The second contribution, described in Section 3, is a modification of Cho et al.'s algorithm which is more resistant against the steganalytic attack. Tests in Section 4 show that the modified algorithm also improves the robustness/capacity trade-off. These two contributions introduce to the field a novel approach to the development of 3D watermarking algorithms, similar to the standard paradigm in image watermarking. That is, the development of algorithms should be evaluated through a competitive process between steganographers and steganalysts.
Limitations: The main limitation of the proposed steganalytic attack is that it has only been tested against the mean-based Cho et al.'s watermarking and its modification proposed in Section 3. As we discuss in Section 5, the steganalytic algorithm could possibly be used against other watermarking methods, however, it might not be straightforward to isolate a statistic with bimodal distribution on marked models.

1.1. Related Work

In the spatial domain, Yeo et al. [3] propose a fragile watermarking method which perturbs a vertex ensuring that predefined hash functions have the same value on it. One of its drawbacks is the causality problem, due to its heavy dependence on the order of traversal of vertices. Lin et al. [4] address this issue using vertex-order-independent hash functions. To increase robustness, Yu et al. [5] and Cho et al. [1], instead of inserting the watermark into a single vertex, embed each watermark bit into a group of vertices. Bors [6] uses a neighborhood localized measure to select the vertices that give small embedding distortion and watermark these vertices by local geometrical perturbations. Aiming at robustness against mesh editing or pose deformation, Yang et al. [7] propose a Laplacian coordinates based algorithm. Steganographic methods include Cayre et al. [8], Wang et al. [9], Chao et al. [10] and Yang et al. [11]. They achieve high capacity and low distortion, but cannot withstand malicious attacks.

In the frequency domain, Ohbuchi et al. [12] propose a method based on the spectral analysis by Karni et al. [13]. It is a non-blind method and thus requires the original mesh during watermark extraction. Using an edge collapse based multiresolution decomposition, Praun et al. [14] propose a robust, non-blind watermarking method. Kanai et al. [15] propose a non-blind method for semiregular meshes based on the modification of wavelet coefficients, while Uccheddu et al. [16] extend this approach to be a blind one.

The area of steganalysis has been primarily developed on images. Fridrich et al. [17] and Ker [18] propose methods specific for the detection of LSB replacement. Farid [19] proposes a universal approach which uses a wavelet-like decomposition to build higher-order statistical models of natural images. Farid's method has been extended in [2] to 3D meshes, which will be used as a benchmark for our approach. Other universal steganalytic approaches for images include Xuan et al. [20], Wang et al. [21] and Lie et al. [22].

2. STEGANALYTIC ALGORITHM

Cho et al.'s mean based algorithm works in the spherical coordinate system. First, a K bin histogram of the radial coordinates is computed and each bin is separately normalised in the interval [0,1]. If $\mathcal{B}'_k = \{\rho_{k,j} : j = 1, 2, 3, ...\}$ is the k-th $(1 \leq k \leq K)$ bin of the normalised radial coordinates $\rho_{k,j}$, a -1 (+1) bit is embedded in that bin by perturbing the vertices such that the mean value $\bar{m}_k$

$$\bar{m}_k = \frac{1}{|\mathcal{B}'_k|} \sum_j \rho_{k,j} \tag{1}$$

is smaller (greater) than 0.5. Here, $|\cdot|$ stands for the number of the elements of a set.

Our steganalytic algorithm is based on the observation that embedding will result in a 2-clustering of the set of the mean values

$$\mathcal{M} = \{\bar{m}_k : 1 \leq k \leq K\} \tag{2}$$

see Fig. 1. The main challenge is finding K, which is done by an exhaustive search through all possible values. For each K, we classify the elements of $\mathcal{M}$ into two clusters using a standard clustering algorithm fitting the data with a mixture of two Gaussians $\mathcal{N}(\mu_{K,i}, \sigma^2_{K,i}), i = 1, 2$. If C and $\tilde{C}$ denote the resulting two clusters, we measure the degree of separation between C and $\tilde{C}$ as the Bhattacharyya distance [23] D_K of the two Gaussians and estimate K by

$$K' = \arg\max_K \{D_K \ : \ K \in [K_{\min}, \ K_{\max}], K \in \mathbb{N}\} \tag{3}$$

subject to

$$\text{abs}\,(|C| - |\tilde{C}|)/K' \leq \epsilon. \tag{4}$$

$K_{\min}$ and $K_{\max}$ define the range of K we would like to consider; we fix $K_{\min} = 30$ and $K_{\max} = 500$.

The rationale behind the constraint of Eq. 4 is the assumption that the watermark bits follow a uniform random distribution, and hence we expect $|C| \approx |\tilde{C}|$. Without the constraint, the distance maximization might return a pair consisting of a small cluster containing a few outliers and a large cluster with all the other values. ϵ in Eq. 4 is a user-specified constant; here $\epsilon = 0.15$.

2.1. Normality Test

After obtaining K', we need to decide whether the mesh has been watermarked or not. We use a normality test, deciding whether $\mathcal{M}$ can be modeled by a single Gaussian, in which case the mesh is clean. Otherwise, we assume that the distribution is bimodal and the mesh has been marked. While standard normality tests exist, here we need a test specifically designed for the extreme cases we deal with. Indeed, since K' is selected for making the distribution of $\mathcal{M}$ as much as possible bimodal, a less sharp test may reject normality even in unmarked meshes.

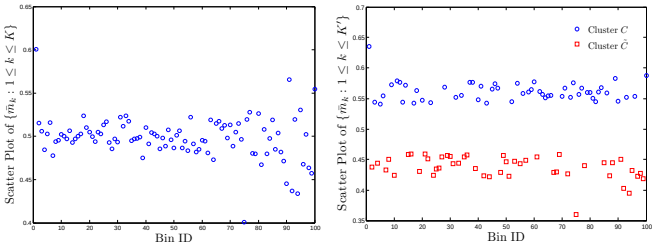

Fig. 1. Scatter plot of the mean values $\{\bar{m}_k : 1 \leq k \leq K\}$ for the clean *Bunny* with $K = 200$ bins (left), and the marked *Bunny* with correct estimation $K' = 200$ bins.

We use Q-Q plots, plotting the quantiles of two distributions against each other. The first distribution is the sample $\mathcal{M}$, while the second is the standardized normal distribution. If two distributions are linearly related, here if $\mathcal{M}$ is linearly related to the normal distribution, the points in the Q-Q plot are nicely modeled by the *reference line* [24]

$$y = \sigma \cdot x + \mu$$

where μ and σ are the mean and the standard deviation of $\mathcal{M}$.

We check if the reference line is a good model of the points of the Q-Q plot by comparing it with the least square linear fit of these points. If the angle θ of the two lines is above a threshold θ_T, we assume that the reference line is not a good model of the Q-Q plot, and hence $\mathcal{M}$ is not Gaussian and the mesh is marked. Fig. 2 shows the Q-Q plots, the reference line and the least square linear fit for the clean and a marked *Rabbit* model. Notice that for reducing the impact of outliers in $\mathcal{M}$, we compute the least square linear fit from the points in the range [-0.5, 0.5] of the normal distribution in the Q-Q plot, the gray area in Fig. 2 (left and middle).

We have also experimented with the most standard approach of applying a t-test on the hypothesis that C and $\tilde{C}$ are independent random samples from normal distributions with equal means. The rejection of the hypothesis of equal means would imply a marked model. However, the results were worse than those from the above tailor-made normality test and we also noticed that one had to use extreme values for the confidence α, e.g., values in the order of $\alpha \approx 10^{-30}$, raising numerical stability concerns.

3. MODIFIED WATERMARKING ALGORITHM

In a bid to increase the robustness of Cho et al.'s algorithm [1] we propose two modifications. First, in the modified algorithm the origin O of the spherical coordinate system is not the barycenter of the vertex set. Instead, we project the vertices onto their principal axis and compute O as the barycenter of the vertices that are projected on that half of the principal axis where the most vertex projections lie. The purpose of shifting O away from the barycenter is to increase the variance in the set of radial coordinates.

The second and most important change is that we embed the message bits by altering the histogram of the radial coordinates (the original method leaves the histogram invariant). While Cho et al.'s approach uses one bin to carry a watermark bit, our method utilizes two bins to deliver a bit. To embed a watermark bit $w_i \in \{-1, +1\}$, we take two neighboring bins of the histogram and possibly transfer some elements from one bin to another. The main idea is that the value of w_i will depend on whether $\mathcal{B}_k$ is shorter or taller than $\mathcal{B}_{k+1}$, that is, on the sign of $|\mathcal{B}_k| - |\mathcal{B}_{k+1}|$.

Starting from the second bin $\mathcal{B}_2$, we arrange adjacent bins into pairs $(\mathcal{B}_2, \mathcal{B}_3), (\mathcal{B}_4, \mathcal{B}_5), \ldots$ and hide a watermark bit into each *embeddable* pair. A pair is embeddable if

$$|\mathcal{B}_k| + |\mathcal{B}_{k+1}| \geq 1 \tag{5}$$

Notice that no watermark bits are carried by the bins $\mathcal{B}_1$ and $\mathcal{B}_K$. That means that the end vertices on the principal axis do not move during embedding, increasing the robustness under blind extraction. If K is odd, the bin pairing process requires to exclude one more bin; here $\mathcal{B}_{K-1}$.

A watermark bit w_i is embedded in an embeddable pair $(\mathcal{B}_k, \mathcal{B}_{k+1})$ by increasing some radial coordinates in $\mathcal{B}_k$, or decreasing some in $\mathcal{B}_{k+1}$. To insert $w_i = -1$, we increase the values of the n_{mov} largest elements ρ_i of $\mathcal{B}_k$, pushing them into $\mathcal{B}_{k+1}$ through

$$\rho_i' = \rho_{\min}^{k+1} + \frac{\Delta_\rho}{\underset{n \in \mathbb{N}, n \geq 3}{\arg\min} \{n : \rho_{\min}^{k+1} + \Delta_\rho / n < \rho_{\max}^{k+1}\}} \tag{6}$$

where ρ_i' is the new radial coordinate, $\rho_{\min}^{k+1}$ and $\rho_{\max}^{k+1}$ are the minimum and the maximum in $\mathcal{B}_{k+1}$ and

$$\Delta_\rho = (\rho_{\max} - \rho_{\min})/K \tag{7}$$

is the range size of each bin. To increase robustness, the denominator of the fraction in Eq. 6 is chosen among a set of possible candidates such that ρ_i' is inside the range of the existing elements of $\mathcal{B}_{k+1}$, that is, $\rho_{\min}^{k+1} < \rho_i' < \rho_{\max}^{k+1}$, and it is as near to $\rho_{\max}^{k+1}$ as possible. Notice that the choice of moving the largest elements of $\mathcal{B}_k$ into $\mathcal{B}_{k+1}$ also helps keeping the embedding distortion to a minimum.

The robustness and distortion trade-off is controlled by a user specified integer threshold $n_{\text{thr}} \geq 1$. We move elements from $\mathcal{B}_k$ into $\mathcal{B}_{k+1}$ until, if possible,

$$|\mathcal{B}_{k+1}'| - |\mathcal{B}_k'| = n_{\text{thr}} \tag{8}$$

We separate the following three cases:
Case 1: If $|\mathcal{B}_{k+1}| - |\mathcal{B}_k| \geq n_{\text{thr}}$, then $n_{\text{mov}} = 0$, meaning no alteration is required.
Case 2: Else if $|\mathcal{B}_k| + |\mathcal{B}_{k+1}| < n_{\text{thr}}$, then $n_{\text{mov}} = |\mathcal{B}_k|$, meaning all the elements in $\mathcal{B}_k$ are transferred into $\mathcal{B}_{k+1}$.
Case 3: Else if $|\mathcal{B}_k| + |\mathcal{B}_{k+1}| >= n_{\text{thr}}$, then

$$n_{\text{mov}} = \left\lceil (|\mathcal{B}_k| - |\mathcal{B}_{k+1}| + n_{\text{thr}})/2 \right\rceil \tag{9}$$

Table 1. Comparison between the universal steganalysis in [2] applied on Cho's et al. (first row), the proposed specific steganalysis applied on Cho's et al. (second row) and the proposed steganalytic algorithm using the statistic $|\mathcal{B}_k| - |\mathcal{B}_{k+1}|$ applied on the proposed modification of Cho's et al. (third row). The fourth column shows the accuracy in the estimation of K and the fifth the steganalytic accuracy.

Method	#Bits	#Marked 3D	Accy of K	Accy
[2]	64	359	N/A	80.93%
Cho's	64	443	96.84%	98.52%
	100	386	96.63%	97.91%
Ours	64	439	87.70%	70.82%
	100	426	92.96%	73.78%

The embedding process for $w_i = +1$ is analogous. It is a straightforward but tedious exercise to show that the above process is reversible and blind extraction of the watermark is possible. The details are omitted.

4. EXPERIMENTAL RESULTS

We validated the steganalytic algorithm on a test set consisting of 445 clean models, mostly from Princeton's University repository [25], and their marked counterparts. As some 3D meshes are unable to carry the watermark for some values of K, we might have different numbers of marked models for different K's.

Table 1 shows that the proposed steganalysis outperforms the universal steganalytic algorithm proposed in [2], while the proposed modified method is more robust. When testing the modified method, we not only applied the steganalytic attack on the distribution of the means $\bar{m}_k$, where it obviously fails (see red curve of Fig. 2 (right)), but also on the differences $|\mathcal{B}_k| - |\mathcal{B}_{k+1}|$ (see green curve of Fig. 2 (right)), trying to detect a possible bimodality on their distribution. Table 1 shows that while by targeting the $|\mathcal{B}_k| - |\mathcal{B}_{k+1}|$ statistic we can achieve a high accuracy rate for the estimation of K, however, the actual rate of steganalytic success is significantly lower. The reason is that the distribution of $|\mathcal{B}_k| - |\mathcal{B}_{k+1}|$ does not follow the single Gaussian assumption as well as that of $\bar{m}_k$.

Fig. 2 (right) plots the detection accuracy with respect to the angle threshold θ_T, measured in degrees, for Cho et al.'s and our watermarking methods that embed 100 bits into the *Rabbit* model. The proposed steganalytic method successfully detects the existence of watermark with an accuracy of up to 98% at $\theta_T = 2.8^\circ$ for Cho et al.'s watermarking and 73% at $\theta_T = 1.5^\circ$ for the modified method. The figure also implies that we can set the threshold θ_T to any value within [2, 3] for Cho et al.'s watermarking and to any value within [1.3, 1.7] for our approach.

Next, we compare the distortion/capacity performance of

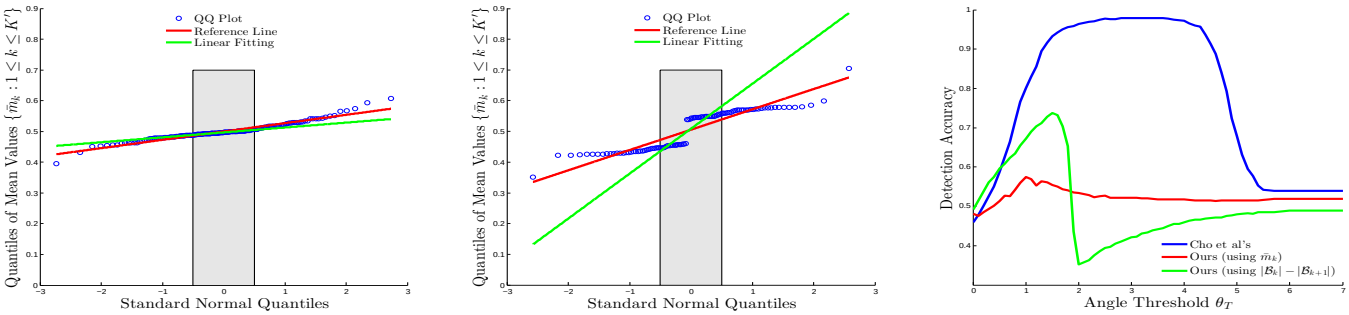

Fig. 2. Q-Q plot of $\mathcal{M}$ for the clean *Rabbit* (left) and the marked by Cho et al.'s with 100 bins (middle). Plot of the detection accuracy with respect to the angle threshold θ_T (in degrees) for Cho et al.'s and the modified method with 100 bits.

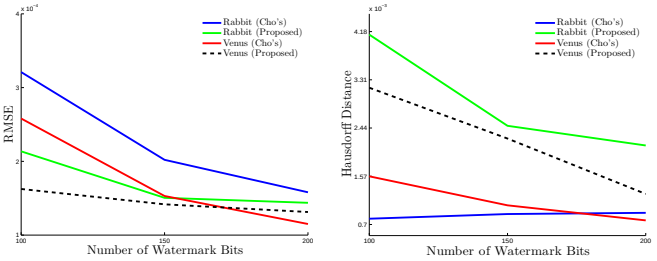

Fig. 3. Embedding distortion measured as the RMSE (left) and the Hausdorff distance (right) when embedding 100, 150 and 200 watermark bits.

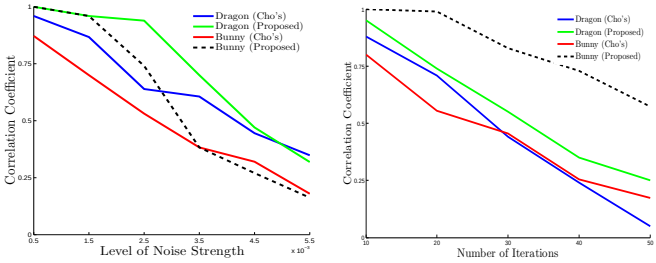

Fig. 4. Comparison between original and modified Cho et al.'s methods. Noise addition (left) and Laplacian smoothing attack (right).

the modified algorithm against the original Cho et al.'s algorithm. Using the Metro tool [26], we measure distortion by the root mean square error (RMSE) and the Hausdorff distance between the clean and the marked mesh. Fig. 3 shows that the performance of the modification is comparable to that of the original.

Finally, we compare the robustness against malicious attacks using the standard measure of the correlation coefficient

$$\mathcal{C}(\mathbf{w}, \mathbf{w}') = \frac{\sum_i (w_i - \bar{w}) \cdot (w'_i - \bar{w}')}{\sqrt{\sum_i (w_i - \bar{w})^2 \cdot \sum_i (w'_i - \bar{w}')^2}} \qquad (10)$$

where $\bar{w}$ and $\bar{w}'$ are the means of the inserted watermark sequence $\mathbf{w}$ and the extracted sequence $\mathbf{w}'$, respectively. Following the 3D mesh watermarking benchmark in [27], we fixed $K = 400$ and carried out attacks with varying strength.
Noise Addition: Random noise was added to all vertex coordinates (x_i, y_i, z_i) according to (resp. y_i, z_i)

$$x'_i = x_i + a_i \cdot \bar{d} \qquad (11)$$

where $\bar{d}$ is the average radial coordinate, and a_i is a uniformly random number in the interval $[-A, A]$. We tested on four different levels of noise : $A = 0.05\%, 0.10\%, 0.25\%, 0.50\%$.
Smoothing: We applied to the marked models 10, 30 and 50 iterations of Laplacian smoothing [28], fixing the deformation factor at $\lambda = 0.02$.

Fig. 4 shows that the modified method is more robust under noise addition and smoothing attacks. In all our experiments, in the original Cho et al.'s method we used the parameter settings recommended in [1].

5. DISCUSSION AND FUTURE WORK

While the proposed steganalytic algorithm was specifically designed to target Cho et al.'s algorithm [1], the main idea could possibly be applied on several other algorithms that embed each watermark bit by altering a specific statistic of the model. For example, a steganalytic attack on the proposed modified algorithm targeting the statistic, $|\mathcal{B}_k| - |\mathcal{B}_{k+1}|$, gives reasonable results. Other algorithms where such a steganalytic attack might be successful include the variance-based watermarking by Cho et al. and the geodesic distance-based watermarking by Luo and Bors [29]. In the future, we plan to adapt the steganalytic algorithm to these cases.

The proposed modification of Cho et al.'s algorithm uses a discrete statistic, here the difference in the height of adjacent bins, rather than a continuous statistic such as the mean of the values in a bin. In the future, we plan to propose discrete counterparts for algorithms such as the one in [29], and check if they also improve the robustness/distortion trade-off, the robustness under steganalytic attacks, and the robustness under malicious watermark removal attacks.

6. REFERENCES

[1] J.-W. Cho, R. Prost, and H.-Y. Jung, "An oblivious watermarking for 3-D polygonal meshes using distribution of vertex norms," *IEEE Trans. on Signal Processing*, vol. 55, no. 1, pp. 142–155, 2007.

[2] Y. Yang and I. Ivrissimtzis, "Mesh discriminative features for 3D steganalysis," *ACM Trans. on Multimedia Computing, Comm & Applications*, vol. 10, no. 3, pp. 27:1–27:13, 2014.

[3] B. L. Yeo and M. M. Yeung, "Watermarking 3D objects for verification," *IEEE Computer Graphics and Applications*, vol. 19, no. 1, pp. 36–45, 1999.

[4] H.-Y.S. Lin, H.-Y.M. Liao, C.-S. Lu, and J.-C. Lin, "Fragile watermarking for authenticating 3-D polygonal meshes," *IEEE Trans. on Multimedia*, vol. 7, no. 6, pp. 997–1006, 2005.

[5] Z. Yu, H. S. Ip, and L. F. Kwok, "A robust watermarking scheme for 3D triangular mesh models," *Pattern Recognition*, vol. 36, no. 11, pp. 2603–2614, 2003.

[6] A. G. Bors, "Watermarking mesh-based representations of 3-D objects using local moments," *IEEE Trans. on Image Processing*, vol. 15, no. 3, pp. 687–701, 2006.

[7] Y. Yang and I. Ivrissimtzis, "Polygonal mesh watermarking using Laplacian coordinates," *Computer Graphics Forum*, vol. 29, no. 5, pp. 1585–1593, 2010.

[8] F. Cayre and B. Macq, "Data hiding on 3-D triangle meshes," *IEEE TSP*, vol. 51, no. 4, pp. 939–949, 2003.

[9] C.-M Wang and Y.-M Cheng, "An efficient information hiding algorithm for polygon models," *Computer Graphics Forum*, vol. 24, no. 3, pp. 591–600, 2005.

[10] M.-W. Chao, C.-H. Lin, C.-W. Yu, and T.-Y. Lee, "A high capacity 3D steganography algorithm," *IEEE TVCG*, vol. 15, no. 2, pp. 274–284, 2009.

[11] Y. Yang, N. Peyerimhoff, and I. Ivrissimtzis, "Linear correlations between spatial and normal noise in triangle meshes," *IEEE TVCG*, vol. 19, no. 1, pp. 45–55, 2013.

[12] R. Ohbuchi, A. Mukaiyama, and S. Takahashi, "A frequency-domain approach to watermarking 3D shapes," *Computer Graphics Forum*, vol. 21, no. 3, pp. 373–382, 2002.

[13] Z. Karni and C. Gotsman, "Spectral compression of mesh geometry," in *SIGGRAPH*, 2000, pp. 279–286.

[14] E. Praun, H. Hoppe, and A. Finkelstein, "Robust mesh watermarking," in *SIGGRAPH*, 1999, pp. 49–56.

[15] S. Kanai, H. Date, and T. Kishinami, "Digital watermarking for 3D polygons using multiresolution wavelet decomposition," in *Proc. Int. Workshop on Geometric Modeling*, 1998, vol. 5, pp. 296–307.

[16] F. Uccheddu, M. Corsini, and M. Barni, "Wavelet-based blind watermarking of 3D models," in *Proc. Workshop on Multimedia and security*, 2004, pp. 143–154.

[17] J. Fridrich, M. Goljan, and R. Du, "Detecting LSB steganography in color, and gray-scale images," *IEEE Multimedia*, vol. 8, no. 4, pp. 22–28, 2001.

[18] A.D. Ker, "Steganalysis of LSB matching in grayscale images," *IEEE SPL*, vol. 12, no. 6, pp. 441–444, 2005.

[19] H. Farid, "Detecting hidden messages using higher-order statistical models," in *ICIP*, 2002, vol. 2, pp. 905–908.

[20] G. Xuan, Y. Shi, J. Gao, D. Zou, C. Yang, Z. Zhang, P. Chai, C. Chen, and W. Chen, "Steganalysis based on multiple features formed by statistical moments of wavelet characteristic functions," in *Proc. Information Hiding Workshop*. Springer, 2005, pp. 262–277.

[21] Y. Wang and P. Moulin, "Optimized feature extraction for learning-based image steganalysis," *IEEE TIFS*, vol. 2, no. 1, pp. 31–45, 2007.

[22] W.-N. Lie and G.-S Lin, "A feature-based classification technique for blind image steganalysis," *IEEE Trans. on Multimedia*, vol. 7, no. 6, pp. 1007–1020, 2005.

[23] E. Choi and C. Lee, "Feature extraction based on the Bhattacharyya distance," *Pattern Recognition*, vol. 36, no. 8, pp. 1703–1709, 2003.

[24] M. Wilk and R. Gnanadesikan, "Probability plotting methods for the analysis for the analysis of data," *Biometrika*, vol. 55, no. 1, pp. 1–17, 1968.

[25] X. Chen, A. Golovinskiy, and T. Funkhouser, "A benchmark for 3D mesh segmentation," in *SIGGRAPH*, 2009, pp. 73:1–73:12.

[26] P. Cignoni, C. Rocchini, and R. Scopigno, "Metro: measuring error on simplified surfaces," *Computer Graphics Forum*, vol. 17, no. 2, pp. 167–174, 1998.

[27] K. Wang, G. Lavoué, F. Denis, A. Baskurt, and X. He, "A benchmark for 3D mesh watermarking," in *Shape Modeling International*. IEEE, 2010, pp. 231–235.

[28] G. Taubin, "Geometric signal processing on polygonal meshes," *Eurographics STAR*, pp. 81–96, 2000.

[29] M. Luo and A. G. Bors, "Surface-preserving robust watermarking of 3-D shapes," *IEEE Trans. on Image Processing*, vol. 20, no. 10, pp. 2813–2826, 2011.