Progressive Randomization: Seeing the Unseen

Anderson Rocha and Siome Goldenstein
Institute of Computing
University of Campinas
13084–851, Campinas, SP – Brazil
{anderson.rocha, siome}@ic.unicamp.br

Abstract

In this paper, we introduce the Progressive Randomization (PR): a new image meta-description approach suitable for different image inference applications such as broad class *Image Categorization* and *Steganalysis*. The main difference among PR and the state-of-the-art algorithms is that it is based on progressive perturbations on pixel values of images. With such perturbations, PR captures the image class separability allowing us to successfully infer high-level information about images. Even when only a limited number of training examples are available, the method still achieves good separability, and its accuracy increases with the size of the training set. We validate the method using two different inference scenarios and four image databases.

Key words: Image Inference; Progressive Randomization; Image Categorization; Hidden Messages Detection; Steganalysis.

1. Introduction

In this paper, we introduce a new image meta-description approach based on information invisible to the naked eye. We validate the new technique on two problems that use supervised learning: *Image Categorization* and *Steganalysis*.

Image Categorization is the body of techniques that distinguish between image classes, pointing out the global semantic type of an image. Here, we want to distinguish the class of an image (e.g., *Indoors* from *Outdoors*). One possible scenario for a consumer application is to group a photo album, automatically, according to classes. Common techniques in content-based image retrieval use color histograms and texture [11], bag of features [20], and shape and layout measures [35] to perform queries in massive image databases. With our solution, we can improve these techniques by automatically restraining the search to one or more classes.

Digital Steganalysis is a categorization problem in which we want to distinguish between *non-stego* or *cover objects*, those that do not contain a hidden message, and *stego-objects*, those that contain a hidden message. Steganalysis is the opposite of *Steganography*: the body of techniques devised to hide the presence of communication. In turn, Steganography is different from *Cryptography*, that aims to make communication unintelligible for those that do not possess the correct access rights. Recently, Steganography has received a lot of attention around the world mainly because its potential applications: identification of sub-components within a data set, captioning, time-stamping, and tamper-proofing (demonstration that original contents have not been altered) [28]. Unfortunately, not all applications are harmless, and there are strong indications that Steganography has been used to spread child pornography pictures on the internet [22]. Robust algorithms to detect the very existence of hidden messages in digital contents can help further forensic and police work. Discovering the content of the hidden message is a much more complex problem than Steganalysis, and involves solving the general problem of breaking a cryptographic code [32].

Here, we introduce the Progressive Randomization (PR): a new image meta-description approach suitable for different image inference applications such as broad class *Image Categorization* and *Steganalysis*. This technique captures statistical properties of the images' LSB channel, information that are invisible to the naked eye. With such perturbations, PR captures the image class separability allowing us to successfully infer high-level information about images.

The PR approach has four stages: (1) the randomization process, that progressively perturbates the LSB value of a selected number of pixels; (2) the selection of feature regions, that makes global descriptors work locally; (3) the statistical descriptors analysis, that finds a set of measure-

ments to describe the image; and (4) the invariance transformation, that allows us to make the descriptor's behavior image independent.

With enough training examples, PR is able to categorize images as a full self-contained classification framework. Even when only a limited number of training examples are available, the method still achieves good separability. The method also provides interesting properties for association with other image descriptors for scene reasoning purposes.

To validate the approach, we use two scenarios: (1) **Broad Image Categorization**; and (2) **Hidden Messages Detection**.

In the **Broad Image Categorization** scenario, we perform four experiments. In the first experiment, we show PR as a complete self-contained multi-class classification procedure. For that, we use a 40,000-image database with 12,000 outdoors, 10,000 indoors, 13,500 art photographs, and 4,500 computer generated images (CGIs) with two different classification approaches: All Pairs majority voting of the binary classifier Bagging of Linear Discriminant Analysis (All-Pairs-BLDA), and SVMs [2]. In addition, we test the PR technique in three other categorization experiments: one to provide another interpretation of the first experiment, one to categorize 3,354 FreeFoto images into nine classes and finally, one to categorize 2,950 fruits images into 15 classes.

In the **Hidden Messages Detection** categorization scenario, we use the 40,000-image database of the first scenario to detect the very existence of hidden messages in digital images. Basically, we want to categorize images into two classes: with and without hidden messages. We use the binary classifiers: Linear Discriminant Analysis with and without Bagging ensemble, and SVMs [2].

We organize the remainder of this paper as follows. Section 2 presents Image Categorization and Steganalysis state-of-the-art. Section 3 introduces the Progressive Randomization approach. Section 4 validates the method for Broad Image Categorization and Steganalysis. Section 5 gives a close study to the reasons of why PR works. Section 6 discusses some method's limitations. Finally, Section 7 draws conclusions and remarks.

2. Related work

In this section, we present recent and important achievements of Image Categorization and Steganalysis.

2.1. *Image Categorization*

Recently, there has been a lot of activity in the area of *Image Categorization*. Previous approaches have considered patterns in color, edge and texture properties to differentiate photographs of real scenes from photographs of art [4]; low- and middle-level features integrated by a Bayesian network to distinguish indoor from outdoor images [17,30]; first- and higher-order wavelet statistics to distinguish photographs from photorealistic images [19].

Fei-Fei et al. [5] have used a Bayesian approach to unsupervised one-shot learning of object categories; Oliva and Torralba [24] have proposed a computational model for scene recognition using perceptual dimensions, coined Spatial Envelope, such as naturalness, openness, roughness, expansion and ruggedness. Bosch et al [3]. have presented an unsupervised scene recognition procedure using probabilistic Latent Semantic Analysis (pLSA). Vogel and Schiele [31] have presented a semantic typicality measure for natural scene categorization.

Recent developments have used middle- and high-level information to improve the low-level features. Li et al. [16] have performed architectonics building recognition using color, orientation, and spatial features of line segments. Some researchers have used bag of features for image categorization [20]. However, these approaches often require complex learning stages and can not be directly used for image retrieval tasks.

2.2. *Digital Steganalysis*

Steganography techniques can be used in medical imagery, advanced data structures designing, document authentication, among others [28]. Unfortunately, not all applications are harmless, and there are strong indications that Steganography has been used to spread child pornography pictures on the internet [22].

In general, steganographic algorithms rely on the replacement of some noise component of a digital object with a pseudo-random secret message [28]. In digital images, the commonest noise component is the Least Significant Bits (LSBs). To the human eye, changes in the value of the LSB are imperceptible, making it an ideal place for hiding information without perceptual change in the cover object.

We can view Steganalysis as a categorization problem in which the main purpose is to collect sufficient statistical evidence about the presence of hidden messages in images, and use them to classify whether or not a given image contains a hidden content.

Westfeld and Pfitzmann [33] have introduced a chi-square-based steganalytic technique that can detect images with secret messages that are embedded in consecutive pixels. Although, their technique is not effective for raw high-color images and for messages that are randomly scattered in the image. Fridrich et al. [7] have developed a detection method based on close pairs of colors created by the embedding process. However, this approach only works when the number of colors in the images is less than 30 percent of the number of pixels. Fridrich et al [8] have analyzed the capacity for lossless data embedding in the least significant bits and how this capacity is altered when a message is embedded. It is not clear how this approach is sensible to different images given that no training stage was applied. Ker [14] has introduced a weighted least-squares

steganalysis technique in order to estimate the amount of payload in a stego object. Notwithstanding, often payload estimators are subject to errors, and their magnitude seem tightly dependent on properties of the analyzed images.

Lyu and Farid [18] have designed a classification technique that decomposes the image into quadrature mirror filters and analyzes the effect of the embedding process.

Fridrich and Pevny [26] have merged Markov and Discrete Cosine Transform features for multi-class steganalysis on JPEG images. Their approach is capable of assigning stego images to six popular steganographic algorithms. Ker [15] have introduced a new benchmark for binary steganalysis based on an asymptotic information about the presence of hidden data. The objective is to provide foundations to improve any detection method. However, there are some issues in computing benchmarks empirically and no definitive answer emerges. Rodriguez and Peterson [29] have presented an investigation of using Expectation Maximization for hidden messages detection. The contribution of their approach is to use a clustering stage to improve detection descriptors.

3. Progressive Randomization approach (PR)

Here, we introduce the Progressive Randomization image meta-description approach for *Image Categorization* and *Steganalysis*. It captures the differences between broad-image classes using the statistical artifacts inserted during the perturbation process.

Algorithm 1 summarizes the four stages of PR: (1) the randomization process; (2) the selection of feature regions; (3) the statistical descriptors analysis; and (4) the invariance transformation.

In the randomization process we progressively perturb the LSB value of a selected number of pixels. Perturbations of different intensities can be carried out. Each one will result a new perturbed image.

With the region selection, we select image regions of interest. For each perturbed image, we select some regions of interest and, for each one, we use statistical descriptors to characterize it.

If we want to evaluate only the relative variations across the perturbations, we perform a normalization with respect to the values in the input image, the one that does not have any perturbation. This amounts to the invariance step.

In summary, if we use $n = 6$ controlled perturbations, we have to analyze the perturbation artifacts in seven images (the input plus the perturbed ones). If we analyze $r = 8$ regions per image and use $m = 2$ statistical descriptors for each region, we have to assess $r \times m = 16$ features for each image. The final PR description vector amounts to $(n+1) \times r \times m = 112$ features. If we perform the last step of invariance (which depends on the application), we normalize each group of features of one perturbed image with respect to the feature values in the input image. The final description vector after normalization amounts to $n \times r \times m = 96$ features.

Algorithm 1 The PR image meta-description approach

Require: Input image I; Percentages $P = \{P_1, \dots P_n\}$;

1: **Randomization:** perform n **LSB pixel disturbances** of the input image ▷ Sec. 3.2

$$\{O_i\}_{i=0\dots n.} = \{I, T(I, P_1), \dots, T(I, P_n)\}.$$

2: **Region selection:** select r feature regions of each image $i \in \{O_i\}_{i=0\dots n}$ ▷ Sec. 3.3

$$\{O_{ij}\}_{\substack{i=0\dots n,\\ j=1\dots r.}} = \{O_{01}, \dots, O_{nr}\}.$$

3: **Statistical descriptors:** calculate m descriptors for each region ▷ Sec. 3.4

$$\{d_{ijk}\} = \{d_k(O_{ij})\}_{\substack{i=0\dots n,\\ j=1\dots r,\\ k=1\dots m.}}$$

4: **Invariance:** normalize the descriptors with respect to the input image I ▷ Sec. 3.5

$$\mathbf{F} = \{f_e\}_{e=1\dots n\times r\times m} = \left\{\frac{d_{ijk}}{d_{0jk}}\right\}_{\substack{i=0\dots n,\\ j=1\dots r,\\ k=1\dots m.}}, \quad (1)$$

5: **Use** $\{d_{ijk}\} \in \mathbb{R}^{(n+1)\times r\times m}$ (non-normalized) or $\{d_{ijk}\} \in \mathbb{R}^{n\times r\times m}$ (normalized) features in your favorite machine learning black box.

3.1. *Pixel perturbation*

Let $\mathbf{x}$ be a Bernoulli distributed random variable with $Prob\{\mathbf{x} = 0\}) = Prob(\{\mathbf{x} = 1\}) = \frac{1}{2}$, B be a sequence of bits composed by independent trials of $\mathbf{x}$, p be a percentage, and S be a random set of pixels of an input image.

Given an input image I of $|I|$ pixels, we define the LSB pixel perturbation $T(I, p)$ the process of substitution of the LSBs of S of size $p \times |I|$ according to the bit sequence B. Let $pixel_i \in S$ be a pixel in S and $b_i \in B$ be an associated bit in B

$$\mathcal{L}(pixel_i) \leftarrow b_i \text{ for all } pixel_i \in S. \quad (2)$$

where $\mathcal{L}(pixel_i)$ is the LSB of $pixel_i$.

Figure 1 shows an example of a perturbation using the bits $B = 1110$.

3.2. *The randomization process*

Given an input image I, the randomization process consists in the progressive application $I, T(I, P_1), \dots, T(I, P_n)$ of LSB pixel disturbances. The process returns n images that only differ in the LSB from the input image, and are identical to the naked eye.

The $T(I, P_i)$ transformations are perturbations of different percentages of the available LSBs. Here, we use $n = 6$

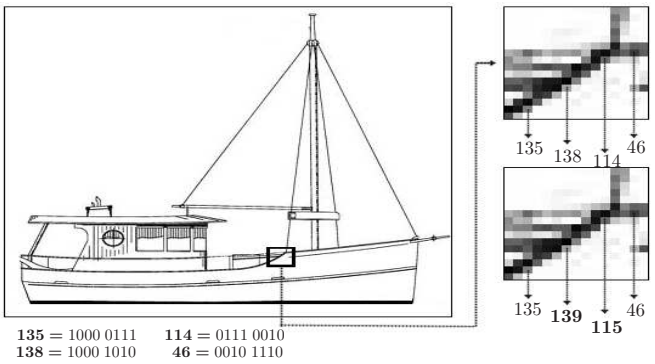


Fig. 1. An example of LSB perturbation using the bits $B = 1110$.

where $P = \{1\%, 5\%, 10\%, 25\%, 50\%, 75\%\}$, $P_i \in P$ denotes the relative sizes of the set of selected pixels S. The greater the LSB pixel disturbance, the greater the randomness of the LSB channel in the resulting image.

3.3. *Feature region selection*

We use statistical descriptors on local regions to capture the changing dynamics of the statistical artifacts inserted during the randomization process (Section 3.2).

Given an image I, we use r regions with size $l \times l$ pixels to produce localized statistical descriptors. In Figure 2, we show the $r = 8$ overlapping regions we use in this paper. We have found out, experimentally, that $r = 8$ regions is a good cost-effectiveness tradeoff.

In [27], we have performed different image feature selection approaches such as identifying regions in the image that are rich in details. To find such regions, we used a filter as defined by Harris and Stephens [12]. In this paper, we have found out, experimentally, that overlapping and non-overlapping regions are enough to capture the scene nuances and no additional regions other than the ones depicted in Figure 2 are used.

3.4. *Statistical descriptors*

The LSB perturbation procedure changes the contents of a selected number of pixels and induces local changes of pixel statistics. An L-bit pixel spans 2^L possible values, and has 2^{L-1} classes of invariance under pixel perturbations (only varies in the LSBs). Let's call these invariant classes *pair of values* (PoV).

When we disturb all the available LSBs in S with a sequence B, the distribution of 0/1 values of a PoV will be the same as in B.

The idea of the statistical analysis is to compare the theoretically expected frequency distribution f_{exp} of the PoVs with the real observed ones f_{obs} [33]. As we know, the perturbation function only affects the LSBs, therefore it does not affect the PoVs distribution after a perturbation. The arithmetical mean remains the same in each PoV, and we can derive the expected frequency through the arithmetic mean between the two frequencies in each PoV.

The observed distribution f_{obs} of PoVs is obtained by counting the PoVs in the image. In turn, the expected frequency f_{exp} is obtained by counting the PoVs after performing the arithmetical mean of their values.

In this paper, we use $m = 2$ statistical descriptors: χ^2 (chi-squared) [33], and U_T (Ueli Maurer) [21]. In Algorithm 1, when we refer to a d_{ijk} descriptor, it is the value of the k^{th} descriptor with $k \in \{U_T, \chi^2\}$ calculated over the j^{th} region $1 \leq j \leq r$, of the i^{th} perturbed image, $0 \leq i \leq n$.

3.4.1. χ^2 *descriptor*

The χ^2 descriptor [6] compares two histograms f^{obs} and f^{exp}. Histogram f^{obs} represents the observations and f^{exp} represents the expected histogram

$$\chi^2 = \sum_{i}^{\nu} \frac{(f_i^{obs} - f_i^{exp})^2}{f_i^{exp}}, \tag{3}$$

where ν is the number of PoVs available. For instance, for RGB images with $L = 8$ bits per pixel, $\nu = 2^{L-1} = 128$.

3.4.2. *Ueli descriptor*

The Ueli descriptor (U_T) [21] is an effective way to evaluate the randomness of a given sequence of numbers. U_T splits an input data S into n blocks. For each block b_i, it analyzes each of the $n-1$ remaining blocks, looks for the most recent occurrence of b_i, and takes the log of the summed temporal occurrences. Let $B(S) = (b_1, b_2, \ldots, b_n)$ be a set of n blocks such that $\cup_{\forall b_i} = S$. We define $U_T : B(S) \rightarrow \mathbb{R}^+$ as a function

$$U_T(B(S)) = \frac{1}{K} \sum_{i=Q}^{Q+K} \ln A(b_i), \tag{4}$$

where K is the analyzed block (e.g., $K = n$), Q is a shift in $B(S)$ (e.g., $Q = \frac{K}{10}$ [21]), and

$$A(b_i) = \begin{cases} i & \nexists i' \in \mathbb{N}, i' < i \rightarrow b_{i'} = b_i, \\ \min\{i' : b_{i'} = b_i\} & \text{otherwise.} \end{cases}$$

According to [21], when evaluated on a sequence of numbers with block size of $|b_i| = 8$ bits, the value of 7.1836 indicates that this sequence most likely is a truly random sequence. On the other hand, the lower the U_T value, the more predictable is the condition in S. In this paper, we consider blocks of size $|b_i| = 8$ given that each component R,G, and B, of an image pixel contains values in the range of $0 \ldots 2^8 - 1$.

3.5. *Invariance*

In some situations, it is necessary to use an image-invariant feature vector. For that, we normalize all descriptors values with regard to their values in the input image (the one with no perturbation)

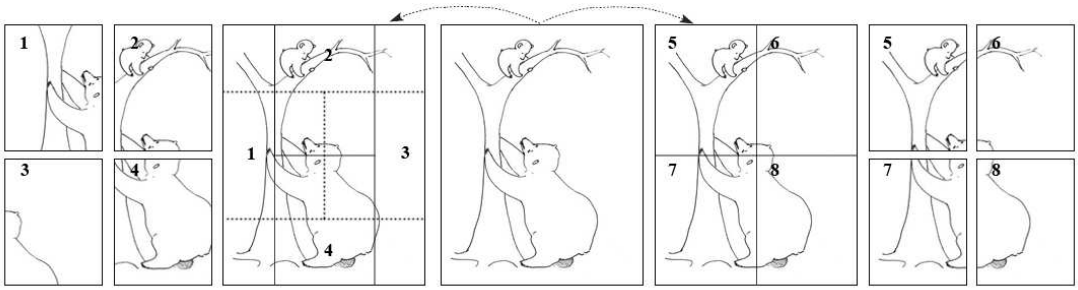


Fig. 2. The eight overlapping regions used in the experiments.

$$F = \{f_e\}_{e=1\ldots n\times r\times m} = \left\{ \frac{d_{ijk}}{d_{0jk}} \right\}_{\substack{i=0\ldots n,\\ j=1\ldots r,\\ k=1\ldots m.}}, \quad (5)$$

where d_{ijk} denotes the k^{th} descriptor $1 \leq k \leq m$ calculated over the j^{th} region $1 \leq j \leq r$ of i^{th} perturbed image $0 \leq i \leq n$. In turn, d_{0jk} runs over the descriptor values calculated on the selected regions of the input image ($i = 0$). F is the final generated descriptor vector of the image I.

Figures 3(a-b) show the behavior of the statistical descriptors along the progressive randomization of one selected image I.

The need for invariance depends on the application. For instance, it is necessary for Steganalysis but harmful for Image Categorization. In Steganalysis, we want to differentiate images that do not contain hidden messages from those that contain hidden messages, and the image class is not important. On the other hand, in Image Categorization, the descriptor values are important to improve the class differentiation.

3.6. *Scalability*

According to the MPEG-7 specification [23], a good image descriptor accounts for optimized efficiency and compactness. PR gives us both. While providing a compact representation, it leads to high classification effectiveness.

MPEG-7 description framework is structured in terms of *Description Schemes* (DS) and *Descriptors* (D), the latter ones instantiated as *Descriptor Values* (DV). In this sense, PR can be used in both ways.

As a description scheme, PR can contain one or more descriptor(s) and/or subordinate description scheme(s) instead of just the χ^2 and U_T values showed here. We could, for instance, perform the controlled perturbations over an image and later on evaluate other common color, and edge descriptors to account their variations.

As a *Descriptor*, we can combine the resulting PR descriptor values with other image descriptors to perform image inference. Indeed, there is room for more research in both directions. Right now we are performing experiments in order to combine other image descriptors under PR description scheme to detect whether or not an image has been manipulated.

4. Experiments and results

In this section, we validate the PR meta-description approach for Broad Image Categorization and Steganalysis. We also compare results with related work in the literature.

For all experiments, we perform K-fold, K = 10, cross-validation in order to provide fair results across the data sets. For that, we partition each data set into K subsets. Of the K subsets, we retain one as the validation data for testing the model, and use the remaining $K - 1$ subsamples as training data. We repeat the process K times (the folds), with each of the K subsets used exactly once as the validation data. Finally, we report average results as well as their standard deviation.

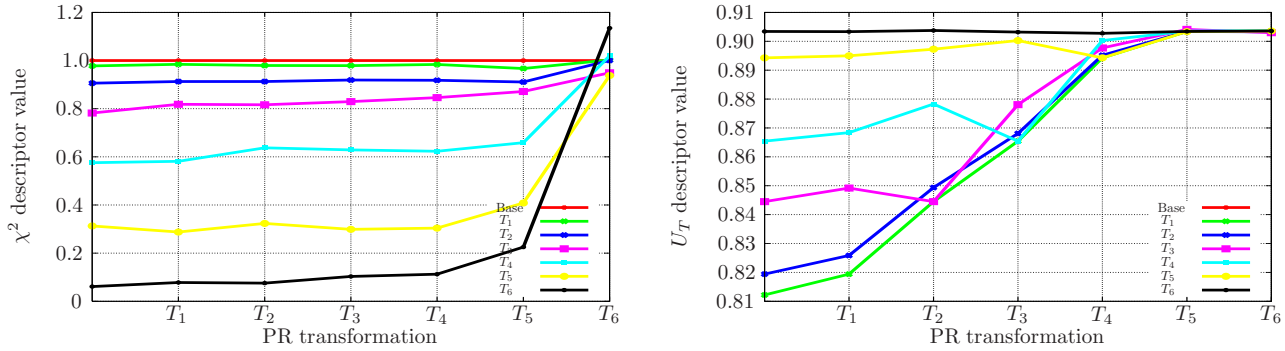
4.1. *Image Categorization*

In this section, we validate the multi-class classification as a complete self-contained classification procedure in **Experiment 1**. In that experiment, we use a 40,000-image database with 12,000 outdoors, 10,000 indoors, 13,500 art photographs, and 4,500 computer generated images (CGIs) with three different classification approaches.

The images in **Experiment 1** come from five main sources: Mark Harden's Artchive [1], the European Web Gallery of Art [2], FreeFoto [3] [4], Berkeley CalPhotos [5], and from The Internet Ray Tracing Competition (IRTC) [6]. Figure 4 shows some examples of each category.

We also validate the PR approach in three other categorization scenarios. In **Experiment 2**, we provide another interpretation of **Experiment 1** using a second carefully assembled image database. In **Experiment 3**, we perform a 9-class image categorization using 3,354 FreeFoto photographs. Finally, **Experiment 4**, we perform a 15-class image categorization using 2,950 images of fruits.

[1] http://www.artchive.com
[2] http://www.wga.hu
[3] http://www.freefoto.com
[4] http://www.ic.unicamp.br/∼rocha/pub/communications.html
[5] http://calphotos.berkeley.edu
[6] http://www.irtc.org



(a) χ^2 descriptor. (b) U_T descriptor.

Fig. 3. Normalized descriptor's behavior along the progressive randomization. T_i represents the PR operation $T(I, P_i)$.

(a) © FreeFoto (b) © Cornelis Saftleven.

(c) (d) © Gilles Tran.

Fig. 4. Examples of each analyzed category: (a) Outdoors, (b) Arts, (c) Indoors, and (d) Computer Generated.

4.1.1. *Experiment 1*

We compare our method to the state-of-the-art two-class separation approaches in the literature [4,19,17,30,25] using a simple *Bagging ensemble* of Linear Discriminant Analysis (BLDA) [9]. Furthermore, we also perform multi-class image-categorization, separating *Outdoor photographs*, *Art images*, *Photorealistic Computer Generated Images*, and *Indoors photographs*.

Two-class classification. Cutzu et al. [4] have addressed the problem of differentiating photographs of real scenes from photographs of art works. They validated over a database with 6,000 photographs from FreeFoto and 6,000 photographs from Mark Harden's Artchive and from Indiana Image Collection [7].

The authors have used color and intensity edges, color variation, saturation, and Gabor features in a complex classifier. We use a similar image set reported in [4]. We have selected 12,000 photographs and 13,500 art photographs totalizing 25,500 images.

[7] http://www.dlib.indiana.edu/collections/dido

Lyu and Farid [19] have used a statistical model based on first- and higher-order wavelet statistics to reveal significant differences of photographs and photorealistic images. They have used photographs from FreeFoto and photorealistic images from IRTC and from Raph 3D Artists.

We use almost the same image set reported in [19]. Therefore, we have used only images from FreeFoto, IRTC and Raph sources, 7,500 photographs and 4,700 photorealistic images, totalizing 12,200.

Luo and Savakis [17,30] have associated texture and color information about sky and grass to differentiate indoors and outdoors images. They have used a Kodak image database not freely available. Payne and Singh [25] have used edge informations to differentiate indoors from outdoors images in a personal image collection.

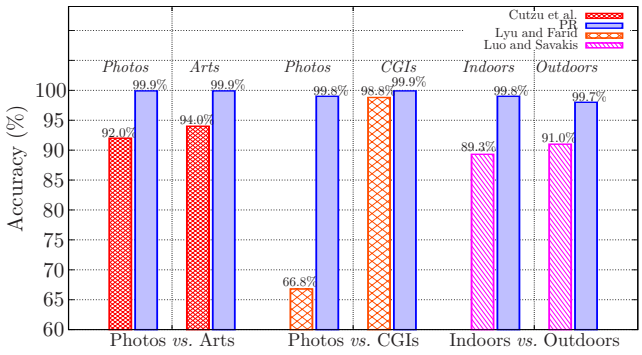


Fig. 5. **Experiment 1**. PR description approach used to binary *Image Categorization* using 10-fold cross-validation.

PR distinguishes *Photographs* from *Art* images with an average accuracy of $\frac{\mu_1+\mu_2}{2} = 99.9.\%$, *photographs* from *CGI* images with an average accuracy of $\frac{\mu_1+\mu_2}{2} = 99.9\%$ and *Indoors* from *Outdoors* images with an average accuracy of $\frac{\mu_1+\mu_2}{2} = 99.7\%$.

Multi-class classification. The PR approach creates a single descriptor that works for different image inference applications. For instance, PR is suitable for multi-class broad image categorization such as the four classes *Indoors*, *Outdoors*, *CGIs*, and *Arts*.

In order to validate the multi-class classification, we have used two different approaches that are combinations of binary classifiers: All Pairs majority voting of the binary classifier BLDA (All-Pairs-BLDA); and Support Vector Machines (SVMs). LibSVM uses an internal mechanism that put together all 1×1 combinations of the classes and performs a majority voting in the final stage. We have used the radial basis function SVM. All-Pairs-BLDA uses sets of binary classifiers and 13 iterations. Note that, any other binary classifier could be used in All-Pairs-BLDA.

Tables 1 and 2 show the resulting classification using All-Pairs-BLDA, and SVMs. The diagonal represents the classification accuracy. For instance, using All-Pairs-BLDA multi-class approach 89.4%, of the images that represent an *Art* scenario are correctly classified, while only 8.17% of them are misclassified as *Indoors*.

The PR approach is independent of the multi-class technique. Figure 6 shows the minimum accuracy of the two approaches as well as the average accuracy and the geometric average accuracy.

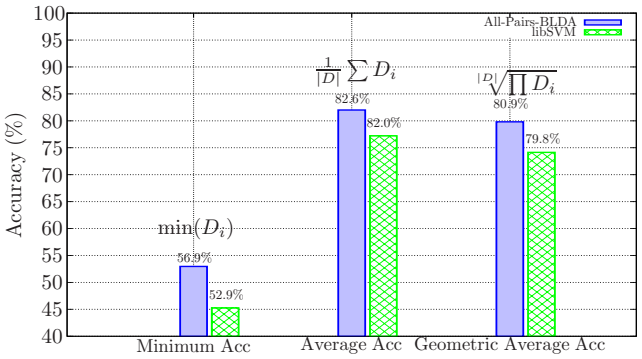


Fig. 6. **Experiment 1**. Multi-class overall accuracy. D is the diagonal of the confusion matrix.

4.1.2. *Experiment 2*

One can argue that the number of training examples in **Experiment 1** is too large and that it might have suffered from bias due to different compression applied to examples of each category, since they come from different sources. It is important to notice that almost all images come from well known image repositories and most of them are built up from user contributions.

We have created a second scenario for multi-class categorization of *Indoors*, *Outdoors*, *CGIs* and *Art photographs*. In this experiment, we have manually selected 500 images for each class totalizing 2,000 images. Each category contains images from at least 75 different internet sources and there are no more than seven images from the same place. There is no intersection among the images in this scenario with the scenario presented in **Experiment 1**. In this experiment, there are at least 400 different cameras with many different compression scenarios. Figure 7 presents the results using the All-Pairs BLDA multi-class approach with 13 iterations.

We show that the results using PR descriptor are not biased due to possible different compression levels. The results are better than the priors of each class (about 25% per class). PR descriptor provides good separation among classes even with few training examples. The more examples we provide in the training phase, the better the classification performance (Figure 7).

4.1.3. *Experiment 3*

Here, we select 3,354 images from FreeFoto and divide them into nine classes. Figure 8 depicts some FreeFoto examples.

We do not pre-process any image. All images come from FreeFoto and were originally stored in JPEG format with 72 DPIs using similar compression levels. Figure 9 presents the results for this experiment using the All-Pairs BLDA multi-class approach with 13 iterations.

Fig. 8. Some FreeFoto categories.

The PR meta-description approach generalizes from the priors (about $\frac{1}{9}$ for each category). The accuracy increases with the number of training examples (left plot of Figure 9). The more images in the training phase, the more accurate the classification. This suggests that PR technique can be combined with other image descriptors for categorization purposes. The average standard deviation σ is bellow 5%. For all classes, the classification results are far above the expected priors (2σ minimum).

4.1.4. *Experiment 4*

In this experiment, we perform categorization of images of fruits [8] and we want to show that the PR results are not biased due to camera properties. We have used the same camera and setup in the capture. The JPEG compression level is the same for all images.

We personally acquired the 2,950 images at the local fruits and vegetables distribution center, using a Canon PowerShot P1 camera, at a resolution of 1024×768 against a white background. Figure 10 depicts the 15 different classes. Even within the same category, there are many illumination differences (Figure 11).

Figure 12 presents the results for this experiment using the All-Pairs BLDA multi-class approach with 13 iterations. Clearly, PR generalizes from the priors (about $\frac{1}{15}$ for

[8] `http://www.ic.unicamp.br/~rocha/pub/communications.html`

	All-Pairs-BLDA Predictions			
	Arts	**CGIs**	**Indoors**	**Outdoors**
Arts	89.4% ± 1.04%	4.41% ± 0.49%	6.16% ± 0.80%	0.00% ± 0.00%
CGIs	33.66% ± 2.36%	53.3% ± 2.09%	13.0% ± 1.22%	0.00% ± 0.00%
Indoors	8.97% ± 0.54%	5.58% ± 0.34%	85.44% ± 0.62%	0.01% ± 0.03%
Outdoors	0.00% ± 0.00%	0.06% ± 0.07%	0.02% ± 0.05%	99.9% ± 0.11%

Table 1
Experiment 1. PR multi-class *Image Categorization* using All-Pairs-BLDA.

	SVM Predictions			
	Arts	**CGIs**	**Indoors**	**Outdoors**
Arts	86.4% ± 1.16%	2.10% ± 0.43%	11.5% ± 0.85%	0.00% ± 0.00%
CGIs	36.2% ± 2.32%	45.3% ± 1.55%	18.2% ± 1.34%	0.20% ± 0.20%
Indoors	17.5% ± 1.28%	4.90% ± 0.41%	77.6% ± 1.47%	0.00% ± 0.00%
Outdoors	0.02% ± 0.03%	0.37% ± 0.18%	0.01% ± 0.03%	99.6% ± 0.21%

Table 2
Experiment 1. PR multi-class *Image Categorization* using SVM.

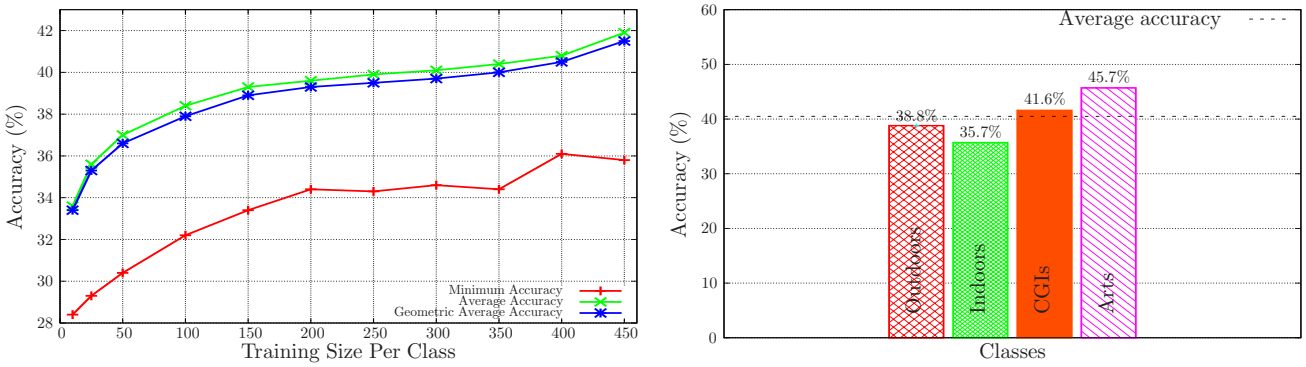


Fig. 7. **Experiment 2**. Out × Ind × CGI × Arts using All-Pairs BLDA(13) .:. 4 classes. *Left plot*: average performance for variable training sizes. *Right plot*: class' performance for 450-sized training sets.

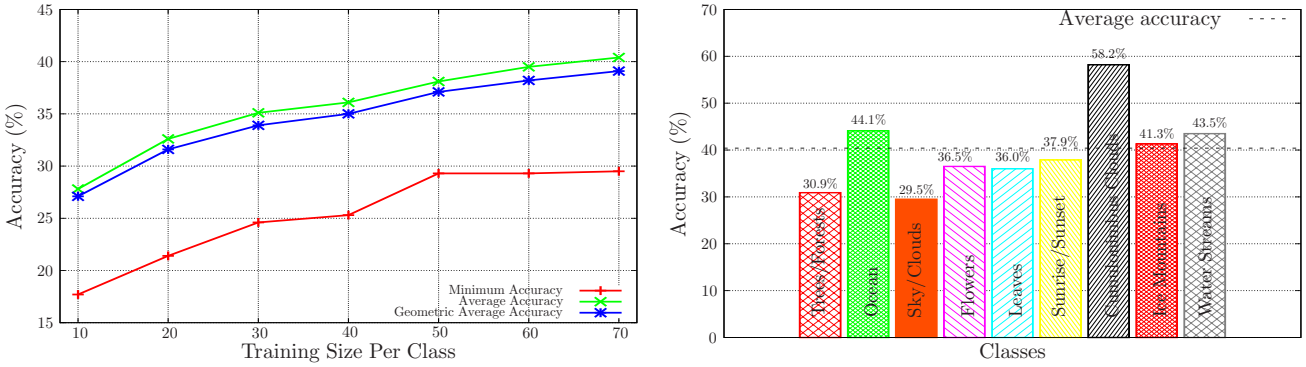


Fig. 9. **Experiment 3**. FreeFoto categorization using All-Pairs BLDA(13) .:. 9 classes. *Left plot*: average performance for variable training sizes. *Right plot*: class' performance for 70-sized training sets.

each category) and the accuracy increases with the number of training examples (left plot of Figure 12).

4.2. *Steganalysis*

In this section, we describe how we train, test and validate PR image meta-description approach for *Steganalysis*. In this scenario, our objective is to detect whether or not a given image contains an embedded content. Here, we have used the same image database of Experiment 1 in Section 4.1.1.

Among all message embedding techniques, the *Least Significant Bit* (LSB) insertion/modification is considered a difficult one to detect [32,28]. In general, it is enough to detect whether a message is hidden in a digital content. For example, law enforcement agencies can track access logs

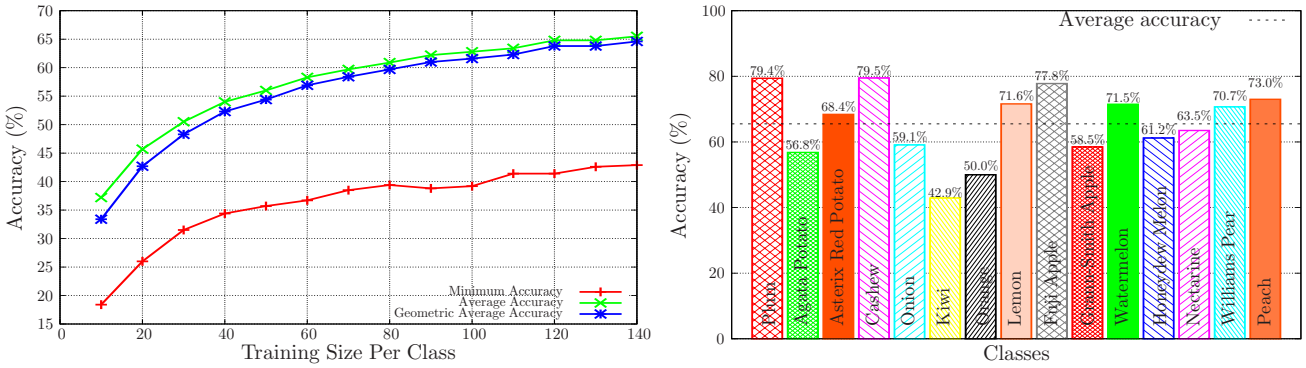


Fig. 12. **Experiment 4**. Fruits categorization using All-Pairs BLDA(13) .:. 15 classes. *Left plot*: average performance for variable training sizes. *Right plot*: class' performance for 140-sized training sets.

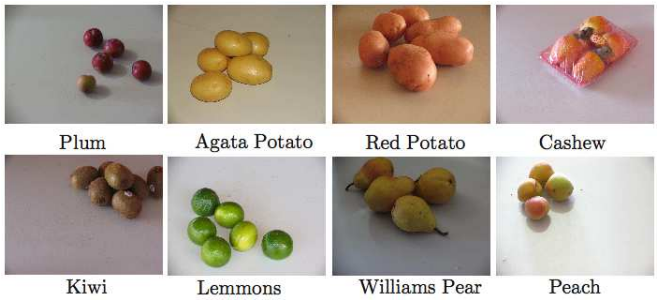


Fig. 10. Some categories of fruits.

Fig. 11. Illumination differences in the *Orange* class.

of hidden contents to build a network graph of suspects. Later, using other techniques, such as physical inspection of apprehended material, they can uncover the actual contents and apprehend the guilty parties [13,28].

4.2.1. *Overall results*

We define a stego-image as an image that suffered an LSB pixel disturbance. The amount of disturbance inserted using the sequence of bits B represents the size of a possible information (message) that is embedded $|M|$. We train a classifier with stego and non-stego examples. To obtain stego examples, we simulate message embeddings perturbing the LSBs of an image subset of our database. We have created a version of our image database for each one of our selected content-hiding scenarios (relative size of contents to the embedding capacity of the image).

In Figure 13, we present the overall results for the PR technique applied for hidden messages detection. We obtain the best results when using the Bagging Ensemble with Linear Discriminant Analysis (BLDA). For instance, for a relative-size message embedding of 10%, PR yields 78.1% of accuracy. That is almost the same result that the more computationally intensive procedure of SVM. Furthermore, it is worth noting that SVM does not benefit from the Bagging ensemble since it is not a weak classifier and it uses only the elements close to the margins in the classification procedure.

PR descriptor scales with the number of examples in the training stage. In overall, the more examples we provide in the training phase, the greater the detection accuracy regardless the message size. In Figure 14, we present the PR descriptor with different training set sizes.

The detection of very small relative-size contents is very hard, and still an open problem. Nevertheless, in practical situations, like when pornographers use images to sell their child-porn images, they usually use a reasonable portion of the LSB channel available space (e.g., 25%). In this class of problem, PR approach detects such activities with accuracy just under 90%.

4.2.2. *Class-based Steganalysis*

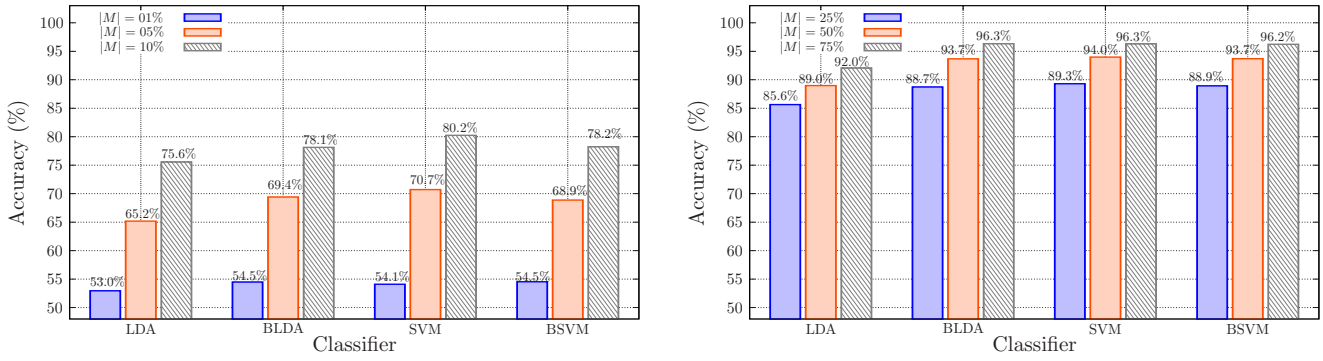
Different classes/categories of images have a very distinct behavior in properties. We explored their different LSB behavior earlier in this paper for proper image categorization.

In this section, we show how the PR descriptor is still able to perform Steganalysis despite all these differences, and gives us a strong insight about which types of images are better for information hiding.

We have found that the detection of hidden content in images with low richness of detail (e.g., *Indoors*) is easier. The inserted artifacts of the embedding process in these images are more obvious than those artifacts inserted in images with more complex details. In these experiment, we have considered four image classes: *Outdoors*, *Indoors*, *Arts*, and *CGIs*. We have used the same image database of Experiment 1 in Section 4.1.1.

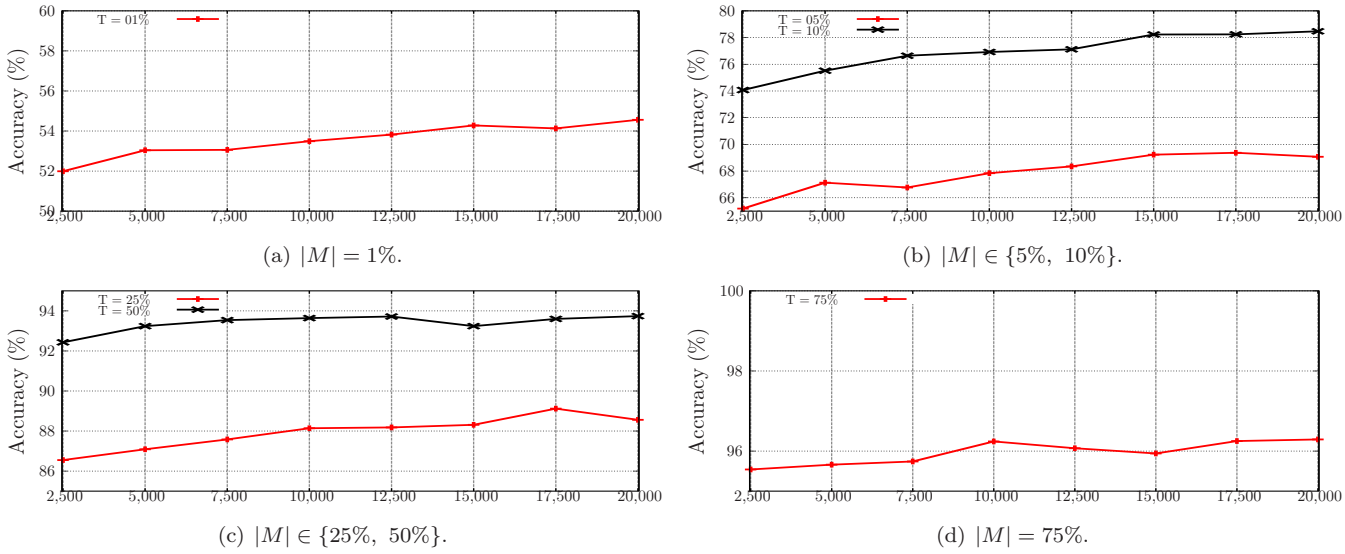
In each analysis, we train our classifier with examples sampled without replacement from three classes, and test in the fourth class. We repeat the process to test each class. Figure 15 shows the resulting classification accuracy for each class of image. We also show the expected classification value when we train and test over all classes with a proportion of 70% examples for training and 30% for testing.

The classes *Arts* and *Outdoors* are the most difficult types to detect hidden messages. On the other hand, *In-*



(a) $|M| \in \{1\%,\ 5\%,\ 10\%\}$ of the LSBs. (b) $|M| \in \{25\%,\ 50\%,\ 75\%\}$ of the LSBs

Fig. 13. PR accuracy for different message embeddings scenarios.



(a) $|M| = 1\%$. (b) $|M| \in \{5\%,\ 10\%\}$.

(c) $|M| \in \{25\%,\ 50\%\}$. (d) $|M| = 75\%$.

Fig. 14. PR accuracy for different training set sizes and stego scenarios.

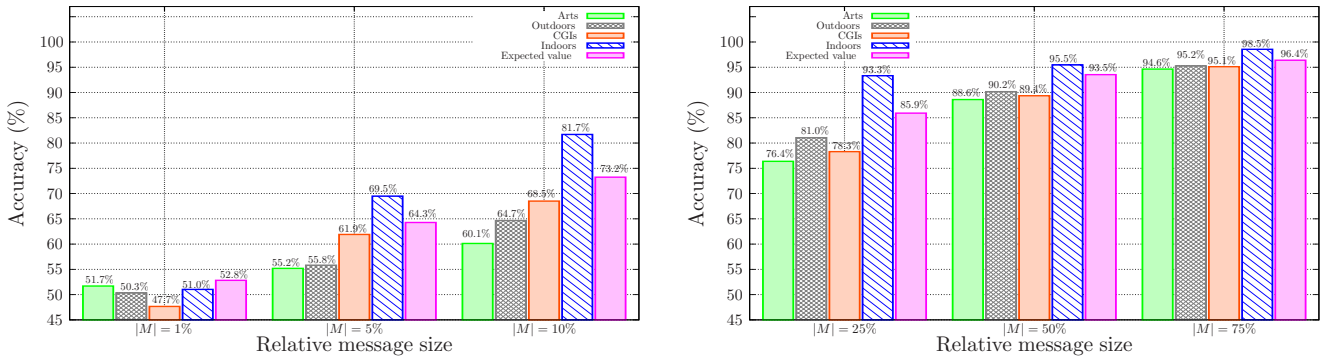


Fig. 15. PR Steganalysis along with different image classes and different relative message sizes.

doors images are the easiest ones. Finally, as our intuition would expect, the greater the message, the better the classification accuracy no matter the class.

4.2.3. *Comparison*

Westfeld and Pfitzmann [33] have devised an approach that only detects sequential hidden messages embedded from the first available LSB. This approach is not robust to image variability and it is not able to detect messages altered by some embedding message procedure that preserves statistics such as mean, and variance about the cover-image. Our framework overcomes these problems and increases the classification accuracy. We compare the results in Table 3. In this experiment, we have used the SVM binary classifier [2].

Other approaches for Steganalysis include the works

of [34,26,10]. However, these approaches have been designed for embedding techniques based on *lossy compression* properties such as those present in JPEG images. In this paper, we present a detection framework designed for *lossless* embedding detection.

We do not expect PR to outperform all Steganalysis algorithms in the literature. Also we are not providing a complete Steganalysis framework. Indeed, in this paper, we present a new image meta-description approach that can be used for Steganalysis. For this reason, we compare our descriptor with two well-known state-of-the-art solutions, and PR association with other image descriptors is straightforward.

	WP	**PR**
	$\mu \pm \sigma$	$\mu \pm \sigma$
01%	$52.6\% \pm 0.1\%$	$54.1\% \pm 0.9\%$
05%	$52.6\% \pm 0.1\%$	$70.7\% \pm 0.9\%$
10%	$54.6\% \pm 4.1\%$	$80.2\% \pm 0.5\%$
25%	$72.9\% \pm 1.9\%$	$89.3\% \pm 0.6\%$
50%	$83.0\% \pm 0.6\%$	$94.0\% \pm 0.5\%$
75%	$84.8\% \pm 0.9\%$	$96.3\% \pm 0.3\%$

Table 3
Westfeld and Pfitzmann's detection approach (WP) *vs.* Progressive Randomization (PR). μ and σ from cross-validation.

Our results are about 26 percentage points better than Westfeld-Pfitzmann's results for small relative-size message embeddings (e.g., $|M| = 10\%$) and are about 17 percentage points better than Westfeld-Pfitzmann's results for medium relative-size message embeddings (e.g., $|M| = 25\%$).

Lyu and Farid [18] have designed a technique that decomposes the image into quadrature mirror filters to analyze the effect of the embedding process. They have used a database of about 40,000 images. The authors tuned their classifiers parameters to have a false positive rate of only 1%. We compare the results in Table 4. The accuracy showed there, for comparison, is the percentage of the stego-images correctly classified. Our Progressive Randomization descriptor detects small (e.g., $|M| = 10\%$) and medium (e.g., $|M| = 50\%$) relative-size message embeddings with an accuracy of about nine percentage points better than Lyu and Farid's approach. When we consider large relative-size message embeddings (e.g., $|M| = 99\%$), our descriptor is about 19 percentage points (about 31 standard deviations) better than Lyu and Farid's approach.

5. Why does PR work?

We initially conceived the Progressive Randomization for Steganalysis of LSB hiding techniques [27]. In this image reasoning scenario, the behavior of the randomization steps is clear: each step emulates hiding a message with a different size. This process is conceptually similar to deciding whether the data is already compressed by looking at the statistics of a new compression operation over this data.

	LDA		**SVM-RBF**		**Type**
	μ	σ	μ	σ	
01%	1.3%	–	1.9%	–	LF
	3.2%	0.5%	3.6%	1.0%	PR
10%	2.8%	–	6.2%	–	LF
	7.0%	0.8%	15.8%	1.1%	PR
50%	16.8%	–	44.7%	–	LF
	24.2%	1.5%	53.1%	1.6%	PR
99%	42.3%	–	78.0%	–	LF
	95.8%	0.5%	97.0%	0.6%	PR

Table 4
Lyu and Farid's detection approach (LF) *vs.* Progressive Randomization (PR) considering FPR = 1%. μ and σ from cross-validation. Lyu and Farid's results from [18].

The experiments in Section 4.1 have showed that PR captures the image class separability allowing us to successfully categorize images. However, the successive-compressions analogy, so intuitive in Steganalysis, is not convincing for this new problem. Our conjecture is that the distinct class behavior comes from the interaction between different light spectrum and the sensors during the acquisition process. That supports the fact that *Outdoors* category is easier to differentiate from the other classes, and that *Indoors* and *Arts* are harder to differentiate amongst each other, as both use artificial illumination.

To show that the separability is not due to different patterns of luminance/color amongst classes, we have devised an experiment to measure the expected value of the U_T descriptor conditioned to the luminance of the region.

We use a local sliding window to calculate local luminance and Ueli, and compute them on all possible 32×32-pixel windows in 300 examples of each class to estimate the $E[U_T|Lum, Class]$, the conditional expectation of U_T given luminance and class.

We approximate the continuous function using histograms of expected values of U_T for each class H_i^E, $i \in \{\text{Outdoor, Indoor}\}$

$$H_i^E \leftarrow E[U_T|L \in 1 \dots 255], \tag{6}$$

where $E[.]$ is the statistical expectation, and L is the luminance such that $L = (0.3 * R) + (0.59 * G) + (0.11 * B)$.

The upper plot of Figure 16 depicts the U_T conditional expectation for unmodified *Outdoors* and *Indoors* classes. There is a consistent difference between classes, showing that the separability of the LSB statistical descriptors is not due to different class patterns of luminance.

We also observe the effect of the limited dynamic range on the statistical descriptor. Luminance components that are too small are forced to zero, while color components that should be very high are pushed to the maximum (255 in the 8-bit case). In these extreme cases, there is no randomness, and the Ueli value goes down to zero. As we calculate the

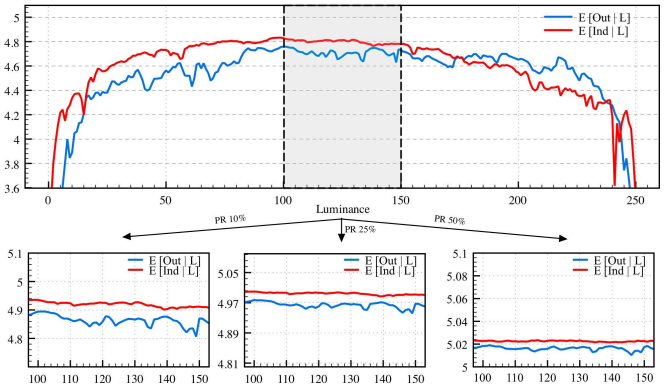


Fig. 16. Dynamic ranges of the conditional U_T descriptor given the luminance variation. From top to bottom, original image set along with perturbations of 10%, 25%, and 50%, respectively.

expected values in sliding windows, the decrease along the borders of the dynamic range demonstrates the decrease of the randomness as more elements of the window are pushed to an extreme value.

6. Limitations

PR technique is not intended to be the final word for Image Categorization and Steganalysis on its own. Of course, there are some limitations with the method. First of all, it is more suitable for loss-less images or high-quality lossy-compressed images. If one performs a medium-to-high-rate lossy compression (e.g., more than 25% of quality loss), it is possible that the method will fail. However, there are a lot of available images in the internet with low-to-medium compression levels. Most of the available cameras in the market use JPEG quality of 85% or more as default (only 15% quality loss due to compression).

Additionally, the approach is sensitive to some kinds of image processing operations. For instance, if the filtering is aggressive enough to destroy the relationship between the LSB channel and the remaining channels, the PR separability will be compromised. For instance, the method stands simple filtering operations such as non-aggressive median and mean, but most likely will fail under more sophisticated ones.

In the case of Steganalysis, if one destroys the LSB channel information using a PRNG, the method potentially will fail. However, in this case, even the message will be destroyed. We report results only for LSB-based Steganography methods. More experiments must be done with respect to the detection of other embedding methods such as JPEG-based ones. This prognostic also applies to filtering. If one performs image filtering, the message will likely be destroyed and it is possible that the method will fail to detect it.

Finally, PR technique probably will fail when used for Image Categorization of images acquired with old cameras with low-quality capturing sensors. In such situations, it is likely that the LSB channel information is related to noise in the process of acquisition. Hence, the relationship of the LSB channel with the other bit channels becomes weaker.

7. Conclusions and remarks

We have introduced a new image descriptor that captures the changing dynamics of the statistical artifacts inserted during a perturbation process in each of the broad-image classes of our interest.

We have applied and validated the Progressive Randomization descriptor in two problems: *Image Categorization* and *Steganalysis*.

The main difference among PR and the state-of-the-art algorithms is that it is based on perturbations on the values of the *Least Significant Bits* of images. With such perturbations, PR captures the image class separability allowing us to successfully infer high-level information about images. Our conjecture is that the interaction of different light spectrum with the camera sensors induces different patterns in the LSB field. PR does not consider semantical information about scenes.

The most important feature in the PR descriptor is its unified approach for different applications (e.g., the class of an image, the class of an object in a restricted domain, hidden messages detection) even with different cameras and illumination.

With enough training examples, PR is able to categorize images as a full self-contained classification framework. However, huge training sets are not always available. When only a limited number of training examples are available, the method still achieves good separability, and its accuracy increases with the size of the training set.

PR uses statistics of the least significant bits for image inference. Although, at first glance, the results might seem surprising, it is not the first time that image bit channels are used for inference. For instance, Avcibas et al. [1] have presented an effective approach for image manipulation detection using statistics across bit channels.

We have demonstrated that PR approach can perform Steganalysis despite differences in the image classes, giving us a strong insight about which types of images are better for information hiding. As our intuition would expect, the greater the message, the better the classification accuracy no matter the class. The detection of very small relative-size contents is very hard, and still an open problem. Nevertheless, in practical situations, like when pornographers use images to sell their child-porn images, they usually have to use a reasonable portion of the LSB channel available space (e.g., 25%). In this class of problem, our approach detects such activities with accuracy just under 90%.

PR descriptor presents two interesting properties that indicate that it can be combined with other image descriptors such as those described earlier in this paper. First, it generalizes from the priors even for small training sets. Second, the accuracy increases with the number of training

examples in all applications we have showed.

Future work includes: to select image regions rich in details and analyze how they are affected using PR descriptor; to investigate other descriptors besides χ^2 and U_T such as kurtosis and skewness; and to apply PR descriptor to other image inference scenarios such as image forgery detection.

Acknowledgments

We thank FAPESP (Grants 05/58103-3, 07/52015-0, and 08/08681-9) and CNPq (Grants 309254/2007-8, 472402/2007-2, and 551007/2007-9) for their support.

References

[1] S. Bayaram, I. Avcibas, B. Sankur, and N. Memon. Image manipulation detection. *Journal of Electronic Imaging (JEI)*, 15(4):1–17, October 2006.

[2] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer Verlab, 2006. ISBN 0-38731-073-8.

[3] A. Bosch, A. Zisserman, and X. Munoz. Scene classification via pLSA. In *ECCV*, 2006.

[4] F. Cutzu, R. Hammoud, and A. Leykin. Distinguishing paintings from photographs. *CVIU*, 100:249–273, 2005.

[5] L. F. Fei, R. Fergus, and P. Perona. One-shot learning of object categories. *TPAMI*, 28(4):594–611, 2006.

[6] D. Freedman, R. Pisani, and R. Purves. *Statistics*. George J. McLeod Limited, 1978.

[7] J. Fridrich, R. Du, and M. Long. Steganalysis of LSB enconding in color images. In *ICME*, volume 3, pages 1279–1282, 2000.

[8] J. Fridrich, M. Goljan, and R. Du. Reliable detection of LSB steganography in color and grayscale images. In *MM&Sec*, pages 27–30, 2001.

[9] J. Friedman, T. Hastie, and R. Tibshirani. *The elements of statistical learning*. Springer Verlag, 2001.

[10] M. Goljan, J. Fridrich, and T. Holotyak. New blind steganalysis and its implications. In *SPIE*, volume 6072, pages 1–13, 2006.

[11] R. Gonzalez and R. Woods. *Digital Image Processing*. Prentice-Hall, 2002.

[12] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Vision Conf.*, pages 147–151, 1988.

[13] N. Johnson and S. Jajodia. Exploring steganography: Seeing the unseen. *IEEE Computer*, 31(2):26–34, 1998.

[14] A. Ker. Optimally weighted least-squares steganalysis. In *SPIE*, volume 6505, 2007.

[15] A. Ker. The ultimate steganalysis benchmark? In *MM&Sec*, pages 141–148, 2007.

[16] Y. Li and L. G. Shapiro. Consistent line clusters for building recognition in cbir. In *ICPR*, volume 3, pages 30952–30957, 2002.

[17] J. Luo and A. Savakis. Indoor vs. outdoor classification of consumer photographs using low-level and semantic features. In *ICIP*, pages 745–748, 2001.

[18] S. Lyu and H. Farid. Detecting hidden messages using higher-order statistics and support vector machines. In *IHW*, pages 340–354, 2002.

[19] S. Lyu and H. Farid. How realistic is photorealistic? *IEEE TSP*, 53:845–850, 2005.

[20] M. Marszałek and C. Schmid. Spatial Weighting for Bag-of-Features. In *CVPR*, pages 2118–2125, 2006.

[21] U. Maurer. A universal statistical test for random bit generators. *Journal of Cryptology*, 5:89–105, 1992.

[22] S. Morris. The future of netcrime now (1) – threats and challenges. Technical Report 62/04, Home Office Crime and Policing Group, 2004.

[23] J.-R. Ohm. The mpeg-7 visual description framework – concepts, accuracy, and applications. In *CAIP*, pages 2–10, London, UK, 2001. Springer-Verlag.

[24] A. Oliva and A. B. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *IJCV*, 42(3):145–175, 2001.

[25] A. Payne and S. Singh. Indoor vs. outdoor scene classification in digital photographs. *Pattern Recognition*, 38(10):1533–1545, 2005.

[26] T. Pevny and J. Fridrich. Merging markov and dct features for multi-class jpeg steganalysis. In *SPIE*, volume 6505, 2007.

[27] A. Rocha and S. Goldenstein. Progressive Randomization for Steganalysis. In *MMSP*, 2006.

[28] A. Rocha and S. Goldenstein. Steganography and steganalysis in digital multimedia: Hype or hallelujah? *Journal of Theoretical and Applied Computing (RITA)*, 15(1):83–110, 2008.

[29] B. Rodriguez and G. Peterson. Steganalysis feature improvement using expectation maximization. In *SPIE*, volume 6575, 2007.

[30] N. Serrano, A. Savakis, and J. Luo. A computationally efficient approach to indoor/outdoor scene classification. In *ICPR*, pages 146–149, 2002.

[31] J. Vogel and B. Schiele. A semantic typicality measure for natural scene categorization. In *DAGM Annual Pattern Recognition Symposium*, 2004.

[32] P. Wayner. *Disappearing cryptography*. Morgan Kaufmann Publishers, 2002.

[33] A. Westfeld and A. Pfitzmann. Attacks on steganographic systems. In *IHW*, pages 61–76, 1999.

[34] Y.Q. Shi et al. Image steganalysis based on moments of characteristic functions using wavelet decomposition, prediction-error image, and neural network. In *ICME*, pages 268–272, Jul 2005.

[35] D. Zhang and G. Lu. Review of shape representation and description. *Pattern Recognition*, 37(1):1–19, 2004.