Computer Vision and Image Understanding 114 (2010) 349–362

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Progressive randomization: Seeing the unseen

Anderson Rocha *, Siome Goldenstein

Institute of Computing, University of Campinas, 13084–851 Campinas, SP, Brazil

ARTICLE INFO

Article history:
Received 13 March 2008
Accepted 21 October 2009
Available online 25 October 2009

Keywords:
Image inference
Progressive randomization
Image categorization
Image forensics
Steganalysis

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, Forensics* 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.

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 subcomponents 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 measurements 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*.

* Corresponding author.
E-mail addresses: anderson.rocha@ic.unicamp.br (A. Rocha), siome@ic.unicamp.br (S. Goldenstein).

1077-3142/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.cviu.2009.10.002

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 4500 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 3354 FreeFoto images into nine classes and finally, one to categorize 2950 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 cannot 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% 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.