18 DÉC

11h00 à 13h00
ENS Paris-Saclay

THÈSES ET HDR

Marina GARDELLA : soutenance de thèse

Titre : Détection automatique et en ligne de la falsification d'images et de vidéos par analyse du bruit.
Direction : M. Colom, J.-M. Morel
Soutenance le 18/12/23 à 11h00 en 1Z34

AJOUTER AU CALENDRIER

Marina GARDELLA

Détection automatique et en ligne de la falsification d'images et de vidéos par analyse du bruit.

Résumé

Les images sont de puissants vecteurs d'information, transmettant une multitude de données et d'informations a travers de représentations visuelles. Leur importance dans divers domaines ne peut être surestimée, car elles offrent des avantages uniques en matière de communication, de compréhension et de documentation. À cette époque, caractérisée par l'influence omniprésente de l'imagerie numérique, la forensique des images représente une discipline vitale qui répond au besoin pressant de maintenir la véracité et la fiabilité du contenu visuel numérique. Les images sont intrinsèquement dotées d'une empreinte digitale, intégrée au cours du processus de formation de l'image. En effet, la création d'une image numérique, depuis son acquisition par le capteur de l'appareil photo jusqu'à son stockage final, imprime des artefacts distincts qui servent de signature unique. L'objectif de cette thèse est de retrouver cette empreinte digitale grâce à l'analyse du bruit. Au long de la chaîne de traitement, le bruit initial de Poisson est transformé par de multiples opérations adaptées à chaque chaîne de formation de l'image, conduisant à l'image comprimée finale. Par conséquent, les résidus de bruit peuvent fournir des informations forensiques significatives.

Tels indices permettent de détecter les falsifications. En effet, bien que les manipulations actuelles permettent d'atteindre un haut degré de fidélité visuelle, elles introduisent simultanément des altérations dans la structure intrinsèque de l'image. Ces perturbations de l'empreinte digitale inhérente sont exploitées par la plupart des méthodes de détection des falsifications pour repérer les régions altérées. La première partie de cette thèse se concentre sur ce problème. Nous proposons ici deux méthodes basées sur la détection des inconsistances locales du modèle de bruit par rapport à un modèle de fond. En particulier, la méthode Noisesniffer adopte une étape de validation a contrario, visant à contrôler le nombre espéré de fausses détections. Nous explorons ensuite la possibilité d'apprendre les traces forensiques en utilisant des réseaux convolutifs profonds, au lieu d'utiliser des features construites à la main. Enfin, cette partie se termine par l'évaluation des méthodes de détection de falsifications elles-mêmes. Nous proposons une méthodologie et un ensemble de données pour étudier la sensibilité des outils de détection à des traces spécifiques, ainsi que leur capacité à effectuer une détection sans indices sémantiques dans l'image.

Les tâches forensiques liées à la caméra source, telles que l'identification du modèle ou la certification du dispositif d'origine, peuvent également être réalisées à l'aide de ladite empreinte digitale. En effet, certaines des traces forensiques intégrées au cours du processus d'acquisition de l'image sont propres au modèle ou à l'appareil. En isolant ces signaux, il est possible d'obtenir des informations sur l'appareil d'origine. La deuxième partie de cette thèse se concentre sur ces tâches. Ici, nous explorons des approches d'apprentissage pour déterminer si une paire d'images contient les mêmes traces forensiques. En outre, nous proposons une nouvelle approche statistique pour la certification de la caméra d'origine basée sur les traces PRNU. Cette approche repose sur deux tests d'hypothèse basés sur des corrélations locales qui ne nécessitent pas le calcul de distributions empiriques.

Cependant, rien n'empêche les faussaires de cacher l'empreinte de l'image. C'est pourquoi nous consacrons la dernière partie de cette thèse à l'analyse de différentes attaques contre-forensiques. Il est important de mettre en évidence les limites des méthodes forensiques actuelles afin de savoir quelle confiance on peut accorder à une image et d'encourager l'exploration d'autres méthodes d'authentification. À cette fin, nous analysons une nouvelle approche récemment introduite dans la littérature pour l'effacement des traces de caméra. Cette approche repose sur une fonction objectif hybride innovante pour l'apprentissage du réseau, définie comme une combinaison de trois fonctions différentes : la fonction de similarité intégrée, la fonction de fidélité tronquée et la fonction d'identité croisée. En outre, nous proposons une nouvelle attaque forensique basée sur des modèles de diffusion.

The avatars of noise in digital images and their use in image forensics

Abstract

Images serve as potent information vectors, conveying a wealth of data and insights through visual representations. Their importance in various domains cannot be overstated, as they offer unique advantages for communication, understanding, and documentation. In an era characterized by the pervasive influence of digital imagery, image forensics represents a vital discipline that addresses the pressing need to uphold the veracity and trustworthiness of digital visual content. Images are naturally endowed with a fingerprint, embedded during the image formation process. Indeed, the creation of a digital image, spanning from its acquisition at the camera sensor to its final storage, imprints distinct artifacts that serve as a unique signature. The goal of this thesis is to retrieve this fingerprint through noise analysis. Along the camera processing pipeline, the initial Poisson noise is transformed by multiple operations tailored to each image formation process, leading to the final compressed image. As a consequence, noise residuals can yield significant forensic insights.

Such cues allow forgery detection. Indeed, though nowadays manipulations have the capability to achieve a high degree of visual fidelity, they concurrently introduce alterations to the intrinsic structure of the image. Such disruptions in the inherent fingerprint are exploited by most forgery detection methods to spot tampered regions. The first part of this thesis focuses on this problem. Here, we propose two methods based on the detection of local inconsistencies of the noise model with respect to a background model. In particular, the Noisesniffer method adopts an a contrario validation step, aiming at controlling the expected number of false detections. We then explore the possibility of learning the forensic traces by means of deep convolutional networks instead of using hand-crafted features. Finally, this part ends with the evaluation of forgery detection methods themselves. We propose a methodology and a dataset to study the sensitivity of the detection tools to specific traces, as well as their ability to perform detection without semantic cues in the image.

Source camera forensics tasks such source camera model identification or source device certification can also be achieved using the said fingerprint. Indeed, some of the forensic traces embedded during the image acquisition process are model-unique or device-unique. By isolating such signals, information about the source device can be obtained. The second part of this thesis focuses on these tasks. Here, we explore learning approaches to determine if a pair of images contain the same forensic traces. In addition, we propose a new statistical approach for source camera certification based on PRNU traces. Such an approach relies on two hypothesis tests based on local correlations which do not require computing empirical distributions.

Still, nothing prevents the forgers from hiding the image fingerprint. This is why we devote the final part of this thesis to the analysis of different counter-forensics attacks. Highlighting the limitations of current forensic methods is important so that one can know how much trust can be put into an image and to encourage the exploration of alternative authentication methods. To this end, we analyze a novel approach recently introduced in the literature for camera trace erasing. This approach relies on an innovative hybrid loss for network training defined as a combination of three different losses: the embedded similarity loss, the truncated fidelity loss and the cross-identity loss. In addition, we propose a new counter-forensic attack based on diffusion models.

Direction

- Miguel COLOM
- Jean-Michel MOREL

Jury

- Agnès Desolneux (CNRS, ENS Paris-Saclay)
- Symeon Papadopoulos (Centre for Research and Technology Hellas)
- Patrick Bas (CNRS, Université de Lille)
- William Puech (CNRS, Université de Montpellier)
- Florent Retraint (Université de Technologie de Troyes)
- Miguel Colom (CNRS, ENS Paris-Saclay)
- Jean-Michel Morel (University of Hong Kong)
- Pablo Musé (Facultad de Ingeniería, UdelaR)