

Marcin Plata

Zastosowanie zaawansowanych metod
sztucznej inteligencji do wybranych
problemów bezpieczeństwa informacji

Rozprawa doktorska

Promotor:

prof. dr hab. inż. Marek Klonowski

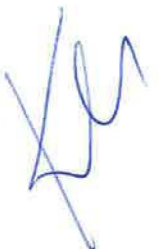
Promotor pomocniczy:

dr inż. Piotr Syga

Instytut Podstaw Informatyki
Polskiej Akademii Nauk

Warszawa 2023

Plata



Streszczenie w języku angielskim (Abstract)

In this dissertation, we present methods categorized as artificial intelligence, which we successfully apply to selected problems in the field of information security. We touch on two main issues – image watermarking and biometrics. Regarding biometrics, we delve into three different topics: standard authentication, spoofing attack robustness, and privacy protection. Each problem is presented based on different biometric features, i.e. hand contour, voice, and facial mimic, respectively.

The difficulty in creating watermarking solutions mainly stems from the need to make methods robust against attacks and compression algorithms such as JPEG. The choice of attacks could significantly affect the solution's architecture required to develop (e.g., the cropping operation disrupts the image spatial structure, significantly hindering the watermark extraction, which is often overlooked in research). We present a new approach based on three neural networks – encoder, decoder, and discriminator, which we extend with our novel method of watermark spatial spreading, which significantly increases robustness against some attacks and JPEG compression compared to recent solutions while keeping high robustness against the cropping attack. We expand the encoder with an additional component that processes the watermark independently of the input image, thus achieving higher robustness and quality without increasing the actual watermark encoding time. Then we classify attacks into groups depending on image processing operations and show that attacks from all groups should be considered in order to achieve high overall robustness. We also note that in real conditions, the content generated with a particular watermarking method occurs relatively rarely, thus increasing the risk of the watermark being wrongly extracted and falsely accusations of piracy. We address this issue by using the discriminator to preliminarily determine the existence of the watermark and we demonstrate the effectiveness of our approach based on proposed metrics.



Plato

Currently, biometric solutions are widely used, including authorization methods for systems that require special precautions, such as banking applications. In this dissertation, we address the issue of privacy protection for biometric methods using facial features. In our solution, we attempt to hide typical static features and process only the movements of characteristic points (facial expressions). Our approach significantly limits the possibility of reconstructing the face from characteristic points, for example in the case of a database leak. To increase privacy protection, we also normalize the features and propose an approach in which we reject the most correlated facial expressions with static features. Our solution is based on several classical machine learning methods, such as support vector machines and random forests. The obtained results allow us to consider our method as weak biometrics, which would also operate as a supporting method for authorization. A use case in which a user speaks a specific phrase for authorization allows our solution to be considered as cancelable biometrics.

Methods for detecting spoofing attacks are crucial elements of biometric systems. One of the most effective attacks and at the same time most difficult to detect is a replay of a recorded voice. We aim to create a method that is able to detect this type of attack. Due to the specificity of the method application, we use small neural networks – LCNN and Bayesian neural network – enabling fast detection based on a few seconds of an audio sample. To train the models, we use our attack-out cross-validation method that takes into account the division into attacks. Additionally, we apply several regularization techniques that enable to train a model that detects attacks non-included in a training set.

In the case of a biometric authorization based on hand contours, we show that using classical image processing methods is satisfactory to create a computationally efficient solution with high accuracy. Our processing pipeline consists of several fast algorithms, such as Gaussian blur, convex hull calculation, and finding the largest triangle inscribing into a contour. Our approach allows the use of a standard and low-cost device that does not require a GPU. Our method accepts samples taken with an office scanner, which additionally allows for the cost reduction of implementing the system.

