

Yunqian Wen · Bo Liu · Li Song ·  
Jingyi Cao · Rong Xie

# Face De-identification: Safeguarding Identities in the Digital Era

 Springer

# Face De-identification: Safeguarding Identities in the Digital Era

Yunqian Wen • Bo Liu • Li Song • Jingyi Cao •  
Rong Xie

# Face De-identification: Safeguarding Identities in the Digital Era

 Springer

Yunqian Wen  
Department of Electronic Engineering  
Shanghai Jiao Tong University  
Shanghai, China

Bo Liu  
School of Computer Science, University of  
Technology Sydney  
Ultimo, NSW, Australia

Li Song  
Department of Electronic Engineering  
Shanghai Jiao Tong University  
Shanghai, China

Jingyi Cao  
Department of Electronic Engineering  
Shanghai Jiao Tong University  
Shanghai, China

Rong Xie  
Department of Electronic Engineering  
Shanghai Jiao Tong University  
Shanghai, China

ISBN 978-3-031-58221-9      ISBN 978-3-031-58222-6 (eBook)  
<https://doi.org/10.1007/978-3-031-58222-6>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG  
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

If disposing of this product, please recycle the paper.

*“In recognition of those committed to safeguarding identities and advancing privacy in the digital realm. Your dedication to the ethical use of technology shapes a future where innovation coexists harmoniously with personal privacy.”*

# Preface

Welcome to “Face De-identification: Safeguarding Identities in the Digital Era.” As the author/editor of this book, I am honored to present this comprehensive exploration into the intricate realm of safeguarding identities in an increasingly digital landscape.

The idea for this book stemmed from a deep-rooted concern for privacy and security in today’s technologically advanced world. The scope of this work encompasses an extensive study of face de-identification techniques, aiming to address the critical challenges faced in protecting identities amid the pervasive use of facial recognition technologies.

Our intent with this book is to offer a thorough examination of various face de-identification methodologies, elucidating their intricacies, strengths, and limitations. Through a structured approach, we have endeavored to present an array of techniques, from obfuscation-based methods to advanced deep generative models, catering to a diverse audience interested in understanding the multifaceted aspects of preserving privacy in digital spaces.

This book is designed for scholars, researchers, practitioners, policymakers, and individuals curious about the intersection of technology and privacy. It serves as a resource for academics delving into the complexities of identity protection, professionals implementing privacy measures, and enthusiasts seeking a deeper understanding of face de-identification in an evolving digital world.

Sydney, NSW, Australia  
November, 2023

Bo Liu

# Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities, the MoE-China Mobile Research Fund Project (MCM20180702) and the National Key R&D Project of China (2019YFB1802701).

# About the Book

Our combined team from University of Technology Sydney (UTS) and Shanghai Jiao Tong University (SJTU) started to work on the topic of face de-identification from 2020. Our findings in state-of-the-art de-identification technology have been invaluable, making the insights and perspectives highly commendable and respected.

In this compelling work, the reader is presented with an insightful journey into the world of face de-identification. As the team's engaging narrative unfolds, you will be guided through the intricate landscape of safeguarding identities in the digital era.

As an expert in privacy protection, I have witnessed the evolution and impact of technologies on our daily lives, especially with respect to privacy and security concerns. The exploration of face recognition and de-identification techniques in this book is timely and essential in our increasingly interconnected world.

This book introduces a comprehensive exploration of face de-identification techniques, shedding light on the complexities and challenges faced in this field. The innovative strategies and ethical considerations presented here mark a significant step forward in the ongoing dialogue on privacy and identity protection. I am confident that this work will contribute significantly to the discourse on privacy and technology, fostering deeper insights and inspiring further advancements in this crucial area.

I commend all my co-authors for their dedication and expertise in compiling this significant contribution. It is my privilege to introduce this impactful work to readers and commend its relevance, timeliness, and scholarly merit.



# Contents

## Part I Introduction

<b>1</b>	<b>Introduction</b> .....	3
1.1	Background and Motivation .....	3
1.2	Face Recognition and Face De-identification .....	4
1.2.1	Face Recognition .....	4
1.2.2	Face De-identification .....	6
1.3	Book Overview .....	8
	References .....	12
<b>2</b>	<b>Facial Recognition Technology and the Privacy Risks</b> .....	15
2.1	Face Recognition Technology .....	15
2.2	Threat Models and Privacy Risks .....	16
2.3	Regulations and Acts on Facial Data Privacy .....	17
2.4	Conclusion and Future Outlook .....	19
	References .....	19

## Part II Face De-identification Techniques

<b>3</b>	<b>Overview of Face De-identification Techniques</b> .....	23
3.1	Face Image De-identification .....	23
3.1.1	Obfuscation-Based Methods .....	23
3.1.2	k-Same Algorithm Based Methods .....	26
3.1.3	Adversarial Perturbation-Based Methods .....	29
3.1.4	Deep Generative Model-Based Methods .....	32
3.2	Face Video De-identification .....	41
3.2.1	Methods of Applying Image De-identification Methods to Videos .....	42
3.2.2	Methods Designed Specifically for Videos .....	43

3.3	Evaluation Metrics	46
3.3.1	Privacy Protection	47
3.3.2	Utility Preservation	48
	References	50
<b>4</b>	<b>Face Image Privacy Protection with Differential Private <math>k</math>-Anonymity</b>	
	<b><math>k</math>-Anonymity</b>	59
4.1	Introduction	59
4.2	Related Works	60
4.2.1	Privacy-Preserving Machine Learning	60
4.2.2	GAN-Based Face Manipulation	61
4.3	Preliminaries	61
4.3.1	Differential Privacy	62
4.3.2	Privacy Amplification	62
4.4	Our Approach	63
4.4.1	Step 1: Attributes Prediction	63
4.4.2	Step 2: Obfuscation	63
4.4.3	Step 3: Image Generation	65
4.5	Experiments	67
4.5.1	Dataset	67
4.5.2	Implementation Details	67
4.5.3	Performance Analysis	67
4.5.4	Quantitative Evaluation	70
4.6	Conclusion	72
	References	72
<b>5</b>	<b>Differential Private Identification Protection for Face Images</b>	75
5.1	Introduction	75
5.2	Related Work	77
5.2.1	Face De-identification Methods Guaranteed by $k$ -Anonymity Theory	78
5.2.2	Face De-identification Methods Guaranteed by $t$ -Closeness Theory	78
5.2.3	Face De-identification Method Guaranteed by Differential Privacy Theory	79
5.3	Preliminaries	81
5.3.1	Problem Formulation	81
5.3.2	Differential Privacy Theory	81
5.3.3	Face Verification and Our Assumptions	83
5.3.4	The Proposed IdentityDP Framework	83
5.3.5	Stage-I: Facial Representations Disentanglement	84
5.3.6	Stage-II: $\epsilon$ -IdentityDP Perturbation	86
5.3.7	Stage-III: Image Reconstruction	86
5.3.8	Training Process	87
5.3.9	Some Discussions About Our Research Topic	88

5.4	Experiments	90
5.4.1	Experimental Setup	90
5.4.2	Evaluation Metrics	90
5.4.3	Implementation Details	91
5.4.4	$\epsilon$ -IdentityDP Mechanism Analysis	91
5.4.5	Comparisons with Traditional Methods	96
5.4.6	Comparisons with SOTA Methods	97
5.4.7	Generalization Ability	102
5.4.8	Computational Overhead	104
5.5	Conclusion and Future Work	104
	References	104
<b>6</b>	<b>Personalized and Invertible Face De-identification</b>	<b>109</b>
6.1	Introduction	109
6.2	Problem Formulation	110
6.3	Our Approach	111
6.3.1	Network Architecture	112
6.3.2	Training Process	113
6.3.3	Protection Process	114
6.3.4	Recovery Process	115
6.4	Experiments	116
6.4.1	Implementation Details	116
6.4.2	Evaluation Results	116
6.5	Conclusion	123
	References	124
<b>7</b>	<b>High Quality Face De-identification with Model Explainability</b>	<b>127</b>
7.1	Introduction	127
7.2	Related Work	130
7.2.1	3D Monocular Face Reconstruction	130
7.2.2	Blind Face Restoration	130
7.3	Methodology	130
7.3.1	Overview of IDEudemon	130
7.3.2	Step I: Parametric Identity Protection	131
7.3.3	Step II: Utility Preservation	132
7.3.4	Loss Function	133
7.4	Experiments	135
7.4.1	Experimental Setup	135
7.4.2	Protective Perturbation Analysis	136
7.4.3	Comparison with SOTA Methods	137
7.4.4	Model Analysis and Ablation Study	140
7.5	Discussion	142
7.6	Conclusion	142
	References	143

**8 Deep Motion Flow Guided Reversible Face Video De-identification** ..... 147

8.1 Introduction ..... 147

8.2 Related Work ..... 150

    8.2.1 Face Video De-identification ..... 150

    8.2.2 Surveillance Video De-identification ..... 152

8.3 Preliminaries of Problem Formulation ..... 153

8.4 Deep Motion Flow Guided Reversible Face Video De-identification ..... 154

    8.4.1 Protection Module ..... 155

    8.4.2 Recovery Module ..... 156

    8.4.3 Motion Flow Module ..... 156

    8.4.4 Affine Transformation Module ..... 157

    8.4.5 The Entire IdentityMask Pipeline ..... 158

8.5 Implementation ..... 160

    8.5.1 Identity Disentanglement Network Configuration ..... 160

    8.5.2 Other Implementation Details ..... 163

8.6 Experiments ..... 164

    8.6.1 Experimental Setup ..... 164

    8.6.2 Comparison in De-identification ..... 165

    8.6.3 Analysis in Identity Recovery ..... 167

    8.6.4 Model Analysis and Discussions ..... 168

8.7 Conclusions ..... 172

References ..... 173

**Part III Conclusion and Future Work**

**9 Future Prospects and Challenges** ..... 179

9.1 Future Prospects and Open Research Questions ..... 179

9.2 Technical Challenges ..... 181

    9.2.1 Low-Complexity and Real-Time De-identification Methods ..... 181

    9.2.2 Preventing Reverse Engineering Attacks of De-identified Faces ..... 181

    9.2.3 Moving Beyond Supervised Learning on Limited Datasets ..... 182

    9.2.4 Multimodal De-identification ..... 182

References ..... 183

**10 Conclusion** ..... 185

**Glossary** ..... 187

# Acronyms

$\epsilon$	Privacy Budget
2D	Two-Dimensional
3D	Three-Dimensional
3DMM	3D Morphable Model
A <sup>3</sup> GAN	Attribute-aware Anonymization Network
AAD	Adaptive Attentional Denormalization
AAM	Active Appearance Model
AE	Auto Encoder
AINet	Attribute-aware Injective Network
AU	Action Unit
BFR	Blind Face Restoration
BIPA	Biometric Information Privacy Act
CA	Coded Aperture
CCPA	California Consumer Privacy Act
cGAN	conditional Generative Adversarial Network
CNN	Convolutional Neural Networks
CS-SFT	Channel-Split Spatial Feature Transform
DGN	Deep Generative Network
DNN	Deep Neural Network
DP	Differential Privacy
DRRDN	Deep Robust Representation Disentanglement Network
FACS	Facial Action Coding System
FATM	Facial Attribute Transfer Model
FID	Fréchet Inception Distance
FIP	Facial Identity-Preserving
GAN	Generative Adversarial Network
GDPR	General Data Protection Regulation
HiSD	Hierarchical Style Disentanglement
ISR	Inverse Super Resolution
JPEG	Joint Photographic Experts Group
LDP	Local Differential Privacy

LPIPS	Learned Perceptual Image Patch Similarity
MAE	Mean Absolute Error
MfM	Multi-factor Modifier
MMDA	Multimodal Discriminant Analysis
NeRF	Neural Radiance Field
OPOM	One Person One Mask
PATE	Private Aggregation of Teacher Ensembles
PCA	Principal Component Analysis
PCC	Pearson's Correlation Coefficient
PDPA	Personal Data Protection Act
PI	Perceptual Indistinguishability
PIPEDA	Personal Information Protection and Electronic Documents Act
PPAS	Privacy-Preserving Attribute Selection
PSNR	Peak Signal-to-Noise Ratio
R <sup>2</sup> VAE	Replacing and Restoring Variational Autoencoder
RMSE	Root-Mean-Square Error
SOTA	State-of-the-Art
SSIM	Structural Similarity Index Measure
SVD	Singular Value Decomposition
VAE	Variational Auto-Encoder
WGAN	Wasserstein Generative Adversarial Network

**Part I**  
**Introduction**

# Chapter 1

## Introduction



### 1.1 Background and Motivation

In recent years, the world has borne witness to a rapid surge in artificial intelligence technologies, particularly those rooted in deep learning, alongside the widespread proliferation of face recognition applications. This technological renaissance, however, brings with it a pressing concern—privacy [1–4]. Amidst these groundbreaking advancements, faces stand out as one of the most sensitive forms of biological information, intimately connected to personal identity. The essence of face recognition lies in its biometric authentication, a characteristic that is both unique and irrevocable. Yet, the consequences of this technology extend far beyond mere identity verification. On the one hand, when harnessed for cross-referencing with other databases, it unveils a wealth of an individual’s sensitive information. A landmark study by Acquisti et al. [5] underscored how faces can serve as the link connecting diverse databases, revealing trails associated with various personas and ultimately undermining privacy. On the other hand, after confirming the identity of a face through face recognition technologies, advanced visual analysis and understanding tools can infer a large amount of sensitive privacy information from the corresponding visual face. For instance, occupation [6] and health status [7]. This poses a serious threat to the security of personal information.

In light of these growing privacy concerns, the field of face de-identification has emerged as a vital research domain within the realms of security and privacy. Face de-identification, a process that conceals facial features while preserving utility for identity-unrelated applications, has found applications in a wide range of scenarios, from anonymizing faces in media interviews and video surveillance [6] to safeguarding privacy in medical research [7], and beyond [8, 9].

The ubiquity of image acquisition in our daily lives—be it sharing personal images on social media, online learning with cameras, or public safety surveillance—renders the need for enhanced privacy protection all the more critical. Existing privacy safeguards often prove inadequate, allowing third parties to collect



human facial images without consent for large-scale data analysis or questionable applications.

Prominent social media platforms like Google, Facebook, and Shutterfly have faced scrutiny for compromising the privacy of millions by inadvertently leaking private photos to commercial entities, thus embroiling themselves in biometric privacy disputes. Conversely, the need for extensive public facial image datasets to fuel the development of cutting-edge deep learning models has led to the creation of invaluable resources. Yet, these repositories carry inherent privacy risks, resulting in increased restrictions on data sharing. Notably, datasets such as Microsoft’s MS-Celeb-1M, Duke’s MTMC, and Stanford’s Brainwash were, at various times, withdrawn from public access due to privacy concerns.

The growing spotlight on privacy issues has prompted the enactment of stringent laws and regulations, notably the General Data Protection Regulation (GDPR) [10, 11], which prohibits companies from collecting, sharing, or analyzing user data without informed consent. Within the GDPR framework, privacy information encompasses “personal data related to an identified or identifiable natural person,” underscoring the paramount importance of protecting personal identity, particularly in the context of facial image data.

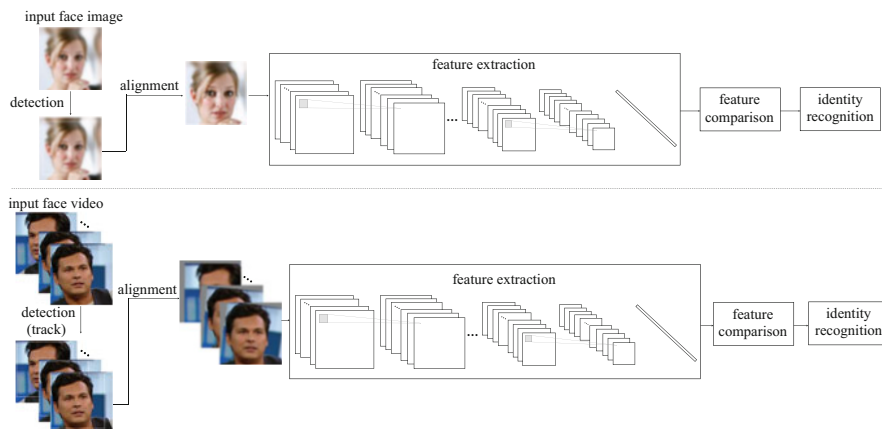
This book, “Face De-identification: Safeguarding Identities in the Digital Era,” endeavors to explore the multifaceted landscape of face de-identification. It delves into a wide array of methods and strategies aimed at preserving facial sensitive information, notably identity, while retaining utility for applications unrelated to identity. Through a comprehensive examination of this crucial field, we seek to provide both practitioners and researchers with the knowledge and tools necessary to navigate the intricate intersection of technology, privacy, and identity protection.

## **1.2 Face Recognition and Face De-identification**

From the background and motivation, it can be seen that face de-identification is a benign technology born to stop face recognition from invading personal privacy, and the two are in a state of confrontation with each other. In order to design excellent face de-identification technology, a thorough understanding of face recognition technology is a necessary condition. Therefore, next we will introduce face recognition and face de-identification separately.

### ***1.2.1 Face Recognition***

Face recognition is a biometric technology that automatically recognizes people’s facial features including statistics and geometric features, which is one of the most important applications of image analysis and understanding. Face recognition tasks can be further divided into binary classification and multiclassification. The binary



**Fig. 1.1** The face recognition process. First, the input image or video is detected and possibly tracks (just for video) to localize the faces. Second, the detected faces are aligned to normalized canonical coordinates. Third, deep facial features are extracted by various methods. After well-designed feature comparison, the identity of the input face data is finally recognized

classification task is also called face verification, which is used to compare whether two images have the same identity. The multiclassification task is also called face retrieval, such as searching for a face with a specific identity in a database of many faces. The widely known face recognition is the abbreviation for identity recognition and verification based on optical facial images. The face recognition process can be simply summarized as using a computer to analyze a face video or image. Firstly, it detects and possibly tracks (just for videos) the faces, so as to localize them. Secondly, it aligns the faces to normalized canonical coordinates. Thirdly, it extracts effective facial features. Finally, it determines the identity of the face object through a comparison of the above-mentioned features. The whole process is shown in Fig. 1.1.

The research on face recognition can be traced back to the late 1960s. The main idea is to design feature extractors and then use machine learning algorithms for classification. Traditional methods rely on hand-made features, such as edge texture description, and combine with machine learning techniques such as principal component analysis, linear discriminant analysis, and support vector machines. The early methods based on geometric features focused on extracting contours and geometric relationships of face components and using the geometric descriptions of shapes and structural relationships as features to construct several feature vectors, including the distance, curvature, and angle between two specified facial keypoints. The advantages are fast recognition speed and low requirements of memory, while the disadvantages are that geometric features can only describe the basic facial information, ignore local subtle features, and result in the loss of local information. The current feature point detection technology is far from meeting the requirements in terms of accuracy.

After introducing deep learning techniques into the field, the approaches have been transferred to extract features with neural networks, which has greatly improved the accuracy and robustness. The deep learning models can be trained by a large amount of data to learn the representation of various variability such as lighting conditions, postures, facial expressions, and so on.

Today, face recognition technology has been widely used in our daily life. Face verification can be treated as a new way of identity confirmation for fast face comparison, mobile payment authentication, security identity verification, etc. Face retrieval can be applied to investigate suspects, complete search of missing persons' databases, and repeated investigation of multiple certificates for one person. At present, the face recognition model can achieve satisfactory accuracy on a specific dataset, but the influence of illumination and posture is still the main challenge. In addition, cross-racial and cross-age recognition problems are also worth studying.

### ***1.2.2 Face De-identification***

Due to potential privacy issues, the application of face recognition technology is currently under controversy, and the face privacy protection task is receiving more and more attention. Face de-identification, the main content of this book, is an innovative technical idea to solve the dilemma. There is no consistent definition of de-identification in the existing literature. Ribaric et al. [12] defined de-identification in multimedia content as *“the process of concealing or removing personal identifiers, or replacing them with surrogate personal identifiers in multimedia content.”* During this process, other facial features that are not related to identity should remain unchanged, such as expression, posture, and background. After this process, the de-identified face will be judged by the face recognition technology as no longer the same identity as the original face. At the same time, the identity-protected face is expected to retain as much similarity to the original image as possible for normal viewing and sharing and can still be analyzed and processed by other identity-agnostic computer vision methods, such as face detection, motion monitoring, and emotion recognition. Additionally, better image quality and visual effects are also preferred.

With face de-identification technologies, visual service providers can use face visual data to carry out legitimate scientific research, business analysis, security monitoring, social sharing, and other activities; ordinary individuals can enjoy the convenience of visual technology without worrying about their other biometric information due to personal identity associated with the disclosure. It effectively alleviates the concerns about personal privacy and security in today's society. To sum up, providing identity protection for facial visual data is the trend of our time, which has great social significance and practical value.

It is recognized that the main purpose of face de-identification is to conceal the identity information of a face. Images and videos are the two main visual data of human faces, and they are also the focus of face de-identification research.