Information Fusion

Jinxing Li • Bob Zhang • David Zhang

Information Fusion

Machine Learning Methods

高等教育出版社
HIGHER EDUCATION PRESS

Springer

Jinxing Li
Harbin Institute of Technology
Shenzhen School of Computer Science and Technology Xili, Nanshan
Shenzhen, China

Bob Zhang
Department of Computer and Information Science
The University of Macau
Macau, China

David Zhang
School of Science and Engineering
The Chinese University of Hong Kong (Shenzhen)
Shenzhen, China

ISBN 978-981-16-8975-8 ISBN 978-981-16-8976-5 (eBook)
https://doi.org/10.1007/978-981-16-8976-5

© Springer Nature Singapore Pte Ltd. & Higher Education Press, China 2022

Jointly published with Higher Education Press, Beijing, China
The print edition is not for sale in China Mainland. Customers from China Mainland please order the print book from: Higher Education Press.
This work is subject to copyright. All rights are reserved by the Publishers, whether the whole or part of the material is concerned, specifically the rights of 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 publishers, 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 publishers nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publishers remain neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

Thanks to the quick development of multimedia techniques, a large amount of data can be collected in different ways. In a large number of real-world applications, the source data or signal is obtained from multiple domains or represented by different measurements. For example, a person could be verified via fingerprint, palmprint, or iris, and a pair of given samples can be measured by the Euclidean or Mahalanobis distance. Generally, the unique representation fails to properly extract valuable information for various applications. Thus, it is significant to conduct research on how to efficiently achieve information fusion.

Information fusion is to combine different types of data to get more accurate, consistent, and useful information compared with any individual modality. Based on the fusion stage, information fusion processes are usually classified into three categories, including low, intermediate, and high levels.

This book introduces various fusion algorithms based on different techniques, including sparse/collaborative representation, Gaussian process latent variable model, hierarchical structure with Bayesian theory, metric learning, score/classifier fusion, and deep learning, respectively. Thanks to these techniques, not only the correlation among multiple features or modalities obtained from a single sample is efficiently exploited but also multiple metrics are utilized to more powerfully and reasonably fit complex scenes of the real-world data. Both do contribute to the performance improvement, and these methods are also widely applied to various pattern recognition fields, including image classification, face verification, disease detection, and image retrieval. The book will be useful for researchers, experts, and postgraduate students, who work in the fields of machine learning, pattern recognition, computer vision, and biometrics. It will be very useful for interdisciplinary research.

Our team has been working on information fusion for a long time. We appreciate the related grant supports from the National Natural Science Foundation of China (NSFC), the China Postdoctoral Science Foundation, the GRF fund of the HKSAR Government, Shenzhen Research Institute of Big Data (SRIBD), and Shenzhen Institute of Artificial Intelligence and Robotics for Society (AIRS).

Shenzhen, China Jinxing Li
Shenzhen, China David Zhang
Macau, China Bob Zhang

Acknowledgments

Our team has been working on information fusion for more than 5 years. We appreciate the related grant supports from the National Natural Science Foundation of China (NSFC) (61906162, 61332011, 61272292, 61271344); Shenzhen Science and Technology Program (RCBS20200714114910193); and the GRF fund of the HKSAR Government, Shenzhen Research Institute of Big Data, and Open Project Fund from Shenzhen Institute of Artificial Intelligence and Robotics for Society (AC01202005017). Besides, we thank Prof. Yong Xu for providing material in Chapter 6.2.

Contents

1 Introduction 1
1.1 Why Do Information Fusion? 1
1.2 Related Works 2
1.2.1 Multi-View Based Fusion Methods 3
1.2.2 Multi-Technique Based Fusion Methods 4
1.3 Book Overview 5
References 8

2 Information Fusion Based on Sparse/Collaborative Representation ... 13
2.1 Motivation and Preliminary 13
2.1.1 Motivation 13
2.1.2 Preliminary 14
2.2 Joint Similar and Specific Learning 16
2.2.1 Problem Formulation 17
2.2.2 Optimization for JSSL 18
2.2.3 The Classification Rule for JSSL 20
2.2.4 Experimental Results 20
2.2.5 Conclusion 26
2.3 Relaxed Collaborative Representation 27
2.3.1 Problem Formulation 27
2.3.2 Optimization for RCR 29
2.3.3 The Classification Rule for RCR 31
2.3.4 Experimental Results 31
2.3.5 Conclusion 36
2.4 Joint Discriminative and Collaborative Representation 37
2.4.1 Problem Formulation 38
2.4.2 Optimization for JDCR 39
2.4.3 The Classification Rule for JDCR 42
2.4.4 Experimental Results 43
2.4.5 Conclusion 47
References 48

3 Information Fusion Based on Gaussian Process Latent Variable Model 51
3.1 Motivation and Preliminary 51
3.1.1 Motivation 51
3.1.2 Preliminary 53
3.2 Shared Auto-encoder Gaussian Process Latent Variable Model 54
3.2.1 Problem Formulation 55
3.2.2 Optimization for SAGP 59
3.2.3 Inference 60
3.2.4 Experimental Results 61
3.2.5 Conclusion 67
3.3 Multi-Kernel Shared Gaussian Process Latent Variable Model 67
3.3.1 Problem Formulation 70
3.3.2 Optimization for MKSGP 74
3.3.3 Inference 76
3.3.4 Experimental Results 77
3.3.5 Conclusion 81
3.4 Shared Linear Encoder-Based Multi-Kernel Gaussian Process Latent Variable Model 81
3.4.1 Problem Formulation 84
3.4.2 Optimization for SLEMKGP 86
3.4.3 Inference 89
3.4.4 Experimental Results 90
3.4.5 Conclusion 97
References 98

4 Information Fusion Based on Multi-View and Multi-Feature Learning 101
4.1 Motivation 101
4.2 Generative Multi-View and Multi-Feature Learning 102
4.2.1 Problem Formulation 103
4.2.2 Optimization for MVMFL 108
4.2.3 Inference for MVMFL 111
4.2.4 Experimental Results 111
4.2.5 Conclusion 115
4.3 Hierarchical Multi-View Multi-Feature Fusion 118
4.3.1 Problem Formulation 118
4.3.2 Optimization for HMMF 121
4.3.3 Inference for HMMF 124
4.3.4 Experimental Results 125
4.3.5 Conclusion 129
References 130

5 Information Fusion Based on Metric Learning 131
5.1 Motivation 131
5.2 Generalized Metric Swarm Learning 133

5.2.1 Problem Formulation 134
5.2.2 Optimization for GMSL 137
5.2.3 Solving Θ with Model Modification 141
5.2.4 Representation of Pairwise Samples in Metric Swarm Space 142
5.2.5 Sample Pair Verification 142
5.2.6 Remarks 144
5.2.7 Experimental Results 146
5.2.8 Conclusion 156
5.3 Combined Distance and Similarity Measure 158
5.3.1 Problem Formulation 158
5.3.2 Optimization for CDSM 162
5.3.3 Kernelized CDSM 164
5.3.4 Experimental Results 166
5.3.5 Conclusion 172
References 172

6 Information Fusion Based on Score/Weight Classifier Fusion 175
6.1 Motivation 175
6.2 Adaptive Weighted Fusion Approach 177
6.2.1 Problem Formulation 177
6.2.2 Rationale and Advantages of AWFA 178
6.2.3 Experimental Results 179
6.2.4 Conclusions 188
6.3 Adaptive Weighted Fusion of Local Kernel Classifiers 188
6.3.1 FaLK-SVM 188
6.3.2 Adaptive Fusion of Local SVM Classifiers 190
6.3.3 Experimental Results 192
6.3.4 Conclusion 194
References 194

7 Information Fusion Based on Deep Learning 197
7.1 Motivation 197
7.2 Dual Asymmetric Deep Hashing Learning 199
7.2.1 Problem Formulation 200
7.2.2 Optimization for DADH 204
7.2.3 Inference for DADH 206
7.2.4 Experimental Results 206
7.2.5 Conclusion 218
7.3 Relaxed Asymmetric Deep Hashing Learning 219
7.3.1 Problem Formulation 222
7.3.2 Optimization for RADH 225
7.3.3 Inference for RADH 228
7.3.4 Implementation 229
7.3.5 Experimental Results 229
7.3.6 Conclusion 235

7.4 Joint Learning of Single-Image and Cross-Image Representations for Person Re-identification 237
7.4.1 Joint SIR and CIR Learning 242
7.4.2 Deep Convolutional Neural Network 245
7.4.3 Experiments 249
7.4.4 Conclusion 253
References 253

8 Conclusion 257

Index 259

List of Figures

Fig. 2.1 The framework of our proposed approach JSSL. There exist three parts in JSSL: dictionary construction, sparse representation, and joint classification. First, the dictionary is formed by training instances. Second, we sparsely represent multi-view data, including sublingual, facial, and tongue features of a given test sample, with the dictionary, and we divide the representation coefficients into two parts, including individual components and similar components. Third, we determine the label of the test instance according to the total reconstruction error 16

Fig. 2.2 Comparison of Healthy and DM performance of the tongue, facial, sublingual features and their fusion methods based on classification. (**a**) Comparison of our method with the tongue image based feature. (**b**) Comparison of our method with the facial image based feature. (**c**) Comparison of our method with the sublingual image based feature 22

Fig. 2.3 ROC curves of different methods and different features for DM detection 24

Fig. 2.4 Classified images based on the tongue, facial, sublingual, and their fusion features for Healthy and DM diagnosis. For each image, the red border indicates incorrect classification, and the green border indicates correct classification 25

Fig. 2.5 Comparison of Healthy vs. IGR performance by using single-modal based methods and JSSL. (**a**) Comparison of our method with tongue image based feature. (**b**) Comparison of our method with facial image based feature. (**c**) Comparison of our method with sublingual image based feature 26

Fig. 2.6 ROC curves of different methods and different features for IGR detection 28
Fig. 2.7 (**a**) The testing instances with sunglasses and scarves in the AR database; (**b**) partitioned testing instances 33
Fig. 2.8 Samples of FRGC 2.0 and LFW. (**a**) and (**b**) are samples in target and query sets of FRGC 2.0; (**c**) and (**d**) are samples in training and testing sets of LFW 33
Fig. 2.9 Illustrations of instances from the Oxford flower datasets. (**a**) Some instances of daffodil in 17 classes; and (**b**) some instances of water lily in 102 classes 35
Fig. 2.10 The framework of the Joint Discriminative and Collaborative Representation model (JDCR). JDRC consists of three steps that are multimodal dictionary construction, Joint Discriminative and Collaborative Representation, and the reconstruction error-based classification. In the first step, training samples belonging to different classes construct the dictionary; multimodal features extracted from the given test instance are then represented by the dictionary, and the shared and discriminative representation coefficient is obtained; third, the label is output by comparing reconstruction errors of each class..... 39
Fig. 2.11 The confusion matrices in 5 independent experiments for Fatty Liver detection. The top left, top right, bottom left, and bottom right corners are sensitivity, false negative rate, false positive rate, and specificity, respectively, followed by their corresponding error bars 46
Fig. 2.12 The ROC curves of various strategies and different features for the Fatty Liver disease detection when the number of training data is 200, 250, and 300 in each category. T-SRC and C-SRC mean the curves are obtained by SRC (tongue) and SRC (Comb), respectively, which are similar to T-DPL and C-DPL..... 47
Fig. 3.1 An overview of the proposed approach. In this model, there is a shared component that can be mapped to different characteristics. In addition, there is another transformation from observations to the shared variable 55
Fig. 3.2 Confusion matrix of the category recognition results using dataset AWA. The vertical axis represents the true labels and the horizontal axis represents the predicted labels 67
Fig. 3.3 Confusion matrix of the category recognition results calculated by JSSL, CCA, DCCA, DCCAE, GPLVM, DS-GPLVM, and SAGP on NUS-WIDE-LITE dataset.

The vertical axis represents the true labels, and the horizontal axis represents the predicted labels 69
Fig. 3.4 The frame of the proposed strategy (MKSGP). We extract multiple features firstly. Then there are mapping functions from the observed samples to the latent space with multiple kernel functions. Meanwhile, mapping functions from the manifold space to the raw data are also computed. Particularly, in order to apply the proposed strategy to the classification, a classifier prior is embedded on the latent component 70
Fig. 3.5 The confusion matrix of category recognition results with dimensionality of the latent variable set 60 using dataset AWA. The vertical axis represents the true labels and the horizontal axis represents the predicted labels 79
Fig. 3.6 The average accuracy of selected 9 categories calculated by various strategies 81
Fig. 3.7 The accuracy calculated by different methods with different values of the dimensionality of the latent variable on the Biomedical dataset 83
Fig. 3.8 The frame of developed strategy SLEMKGP. We first extract multiple views or features. Then, there is a projection for each view, which is learned to map the observations to a consistent space by transformation matrices $\{\mathbf{P}^v\}_{v=1}^V$. A Gaussian process prior is subsequently utilized to obtain a back project to the shared and latent manifold space. Later, the transformations from the shared and latent space to the observed space are also calculated. To use developed strategy SLEMKGP for classifying, a discriminative prior is embedded on the latent component. It should be noted that, $\mathbf{Y}^v$ refers to the input in the v-th view, $\mathbf{P}^v$ refers to associating mapping matrix, and f^v refers to the projection function embedding the GP prior at the v-th view. Here the covariance matrices from both encoder and decoder are computed by utilizing multiple kernels 83
Fig. 3.9 The confusion matrix calculated by SLEMKGP with q set as 50 using AWA dataset. And the elements on the diagonal representing classifying results 94
Fig. 3.10 The average accuracy values from selected 9 classes calculated by JSSL, CCA, DCCA, DCCAE, GPLVM, DSGPLVMI, DSGPLVMS, and SLEMKGP 96
Fig. 4.1 The example containing multiple views and multiple features 102
Fig. 4.2 The framework of the proposed method, where the green round means the observed features, the blue round means

the latent variable, and the orange round means the observed label; J means the number of views and K_j means K_j types of features are obtained in j-th view 103
Fig. 4.3 The probabilistic framework and graphic model of the presented approach. (**a**) The probabilistic framework of the presented approach. (**b**) The graphical model of the presented approach. $\mathbf{x}_{ijk_j}$ and z_{pi} are the observed features and labels, respectively. $\mathbf{h}_{ijk_j}$ is the latent variable associated with $\mathbf{x}_{ijk_j}$. For the label z_{pi}, it follows the multinomial distribution, where $z_{pi} \in \{0, 1\}$ and $\sum_{p=1}^{P} z_{pi} = 1$. According to the supervised information, the latent variable $\mathbf{h}_{ijk_j}$ is generated by following a Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}_{jp}, \boldsymbol{\Sigma}_{jp})$. Then we also assume that $\mathbf{x}_{ijk_j}$ is generated from its associated latent variable, and $\mathbf{x}_{ijk_j}$ also follows a Gaussian distribution $\mathcal{N}(\mathbf{A}_{jk_j}\mathbf{h}_{ijk_j}, \boldsymbol{\Sigma}_{jk_j})$. (**c**) The graphical model of the transformed class-conditional structure of the proposed method. By integrating out the latent variable $\mathbf{h}_{ijk_j}$, all parameters $\{z_{pi}, \boldsymbol{\mu}_{jp}, \boldsymbol{\Sigma}_{jp}, \mathbf{A}_{jk_j}, \boldsymbol{\Sigma}_{jk_j}\}$ are directly imposed on $\mathbf{x}_{ijk_j}$, and $\mathbf{x}_{ijk_j}$ follows a novel Gaussian distribution $\mathcal{N}(\mathbf{A}_{jk_j}(\sum_{p=1}^{P} z_{pi}\boldsymbol{\mu}_{jp}), \boldsymbol{\Sigma}_{jk_j} + \mathbf{A}_{jk_j}(\sum_{p=1}^{P} z_{pi}\boldsymbol{\Sigma}_{jp})\mathbf{A}_{jk_j}^{T})$ 105
Fig. 4.4 The visualization of the generated data when $J = 3$ and $K_j = 4$ 112
Fig. 4.5 The visualization of the original and generated data. (**a**) The original data. (**b**) The generated data 114
Fig. 4.6 ROC curves of different approaches for DM diagnosis 117
Fig. 4.7 (**a**) The framework of the proposed method, where the number of the latent variables is equal to the number of observed views. (**b**) The probabilistic framework of the proposed method 119
Fig. 4.8 The comparison of the distributions between the original data and generated data 127
Fig. 4.9 The ROC curves obtained by different methods in DM detection 129
Fig. 5.1 Illustration of our GMSL for face verification 134
Fig. 5.2 A diagram of the proposed GMSL for classification. Left: the circles with different colors denote similar pairs, the squares with different colors denote dissimilar pairs; medium: the triangles denote the represented vector space of similar pairs, and the diamonds denote the represented vector space of dissimilar pairs; right: the binary classification in the learned metric space. (For interpretation of the references to color in this figure

legend, the reader is referred to the web version of this article) 143
Fig. 5.3 Duality gap vs. the iterations on LFW data 145
Fig. 5.4 Classification error (%) of Euclidean, DML-eig, LMNN, ITML, Sparse ML, Sub-SML, and the proposed GMSL metric learning methods 147
Fig. 5.5 Average rank of the classification error rate 148
Fig. 5.6 The weights of MetricFusion for different q-values (a) and performance curve of three features with different q-values ($q = 2, \ldots, 10$) 152
Fig. 5.7 Comparisons of ROC curves and AUCs between our GMSL and the state-of-the-art methods on LFW 154
Fig. 5.8 Comparison of ROC curves and AUCs 155
Fig. 5.9 Comparison with single metric and combination of four metrics. 1: LBP; 2: SIFT; 3: Attribute; 4: LBP + SIFT; 5: LBP + Attribute; 6: SIFT + Attribute; and 7: LBP + SIFT + Attribute 156
Fig. 5.10 Some intra and inter pairs of faces in PubFig 156
Fig. 5.11 ROC curves and AUCs of Sub-SML and GMSL 157
Fig. 6.1 Steps 3 and 4 180
Fig. 6.2 Some visible and infrared face images. The first row shows the visible face images. The second row shows the infrared face images 181
Fig. 6.3 Some 2D palmprint images from the 2D plus 3D palmprint dataset 182
Fig. 6.4 Four ROI images of a same palm. The first, second, third, and fourth ROI images were extracted from the red, green, blue, and near infrared images, respectively 184
Fig. 6.5 Some image samples from the GT face dataset 185
Fig. 6.6 Sample images of one individual from the LFW dataset 187
Fig. 7.1 The framework of the proposed method. Two streams with five convolution layers and three full-connected layers are used for feature extraction. For the real-valued outputs from these two neural networks, their similarity is preserved by using a pairwise loss. Based on the outputs, a consistent hash code is generated. Furthermore, an asymmetric loss is introduced to exploit the semantic information between the binary code and real-valued data 201
Fig. 7.2 The Precision-Recall curves computed by LSH, SpH, ITQ, SDH, SGH, DPSH, ADSH, DAPH, and DADH on the IAPR TC-12 dataset. Figures from (**a**) to (**f**) are associated with the code length of 8-bit, 12-bit, 16-bit, 24-bit, 36-bit, and 48-bit 210
Fig. 7.3 The Precision-Recall curves computed by LSH, SpH, ITQ, SDH, SGH, DPSH, ADSH, DAPH, and DADH on

the MIRFLICKR-25K dataset. Figures from (**a**) to (**f**) are associated with the code length of 8-bit, 12-bit, 16-bit, 24-bit, 36-bit, and 48-bit 213

Fig. 7.4 The Precision-Recall curves computed by LSH, SpH, ITQ, SDH, SGH, DPSH, ADSH, DAPH, and DADH on the CIFAR-10 dataset. Figures from (**a**) to (**f**) are associated with the code length of 8-bit, 12-bit, 16-bit, 24-bit, 36-bit, and 48-bit 216

Fig. 7.5 The changing of the objective function values and MAP scores with an increase in iterations 217

Fig. 7.6 The changing of objective function values and MAP scores with an increase in iterations. Note that the asymmetric terms $\left\|\tanh(\mathbf{F})\mathbf{B}^T - k\mathbf{S}\right\|_F^2$ and $\left\|\tanh(\mathbf{G})\mathbf{B}^T - k\mathbf{S}\right\|_F^2$ are removed 218

Fig. 7.7 The motivation of the point-to-angle matching. Assume $\mathbf{f}_1$, $\mathbf{f}_2$, $\mathbf{f}_3$, $\mathbf{f}_4$, $\mathbf{f}_5$, $\mathbf{f}_6$, and $\mathbf{b}_1$ belong to the same Hamming space $\mathbf{R}_1$. (1) $(\mathbf{f}_1^T\mathbf{f}_2 - k)^2$ makes them similar and discrete. However, if $\mathbf{f}_1$ and $\mathbf{f}_2$ are both transformed to $\mathbf{f}_1'$ and $\mathbf{f}_2'$ in $\mathbf{R}_2$, the value of $\mathbf{f}_1^T\mathbf{f}_2$ remains unchanged but their hashing codes change. (2) The asymmetric inner product $(\mathbf{f}_1^T\mathbf{b}_1 - k)^2$ can well tackle the problem in (1), but this term strongly enforces each element in $\mathbf{f}_1$ ($\mathbf{f}_2$) to approximate to be either $+1$ or -1. This is too strict and unnecessary since what we need is $\text{sign}(\mathbf{f}_1)$. (3) Thus, some works try to use the cosine distance to measure the similarity between each pair of samples. Although this strategy can address the existing problem in (2), it may be unable to process some specific cases like $(\mathbf{f}_3, \mathbf{f}_4)$. (4) In order to solve the problems mentioned in (1), (2), and (3), we propose a relaxed asymmetric strategy that exploits the cosine distance between the real-valued outputs and the binary codes. In this figure, $(\frac{\langle \mathbf{f}_5, \mathbf{b}_1 \rangle}{\|\mathbf{f}_5\|\|\mathbf{b}_1\|} - 1)^2$ or $(\frac{\langle \mathbf{f}_6, \mathbf{b}_1 \rangle}{\|\mathbf{f}_6\|\|\mathbf{b}_1\|} - 1)^2$ not only encourages them onto the same hypercube without any length constraint but also efficiently avoids the case that occurs between $\mathbf{f}_3$ and $\mathbf{f}_4$ 220

Fig. 7.8 The motivation of the novel triplet loss. $\mathbf{f}_1$ and $\mathbf{f}_2$ belong to the same class, while $\mathbf{f}_3$ belongs to another class. According to the analysis in Fig. 7.7, $\mathbf{f}_1$ and $\mathbf{f}_2$ are located in appropriate positions. However, the Euclidean distance between them is much larger than that of $\mathbf{f}_2$ and $\mathbf{f}_3$. The traditional triplet cannot be directly used. Thus, we propose a novel triplet loss. Firstly, each output is normalized onto a unit ball to get $\bar{\mathbf{f}}_1$, $\bar{\mathbf{f}}_2$, and $\bar{\mathbf{f}}_3$ corresponding to $\mathbf{f}_1$, $\mathbf{f}_2$, and $\mathbf{f}_3$, respectively. Then

$[1 - (\bar{\mathbf{f}}_1^T \bar{\mathbf{f}}_3 - 1)_2^2 + (\bar{\mathbf{f}}_1^T \bar{\mathbf{f}}_2 - 1)_2^2]_+$ is constructed to encourage $\bar{\mathbf{f}}_1$ to be closer to $\bar{\mathbf{f}}_2$ than to $\bar{\mathbf{f}}_3$ under the Hamming distance 221

Fig. 7.9 The framework of the proposed method. The top and bottom streams with different weights are used to perform the feature extraction for the images. A shared binary code $\mathbf{b}_i$ is then learned corresponding to the same input in both streams. Here we assume that $(\mathbf{b}_1, \mathbf{f}_1, \mathbf{f}_2)$ have the same semantic information, while $(\mathbf{b}_2, \mathbf{f}_3)$ enjoy different semantic information from $\mathbf{b}_1$. In the learning step, the relaxed asymmetric is exploited to make $(\mathbf{b}_1, \mathbf{f}_1, \mathbf{f}_2)$ locate in the same Hamming space without the length constraint. We also propose a novel triplet loss to rank the positive and negative pairs 222

Fig. 7.10 The Precision-Recall curves computed by LSH, SpH, ITQ, SDH, SGH, DPSH, ADSH, DAPH, and RADH on the CIFAR-10 dataset. Figures from (**a**) to (**f**) are associated with the code length of 8-bit, 12-bit, 16-bit, 24-bit, 36-bit, and 48-bit 231

Fig. 7.11 The Precision-Recall curves computed by LSH, SpH, ITQ, SDH, SGH, DPSH, ADSH, DAPH, and RADH on the NUS-WIDE dataset. Figures from (**a**) to (**f**) are associated with the code length of 8-bit, 12-bit, 16-bit, 24-bit, 36-bit, and 48-bit 233

Fig. 7.12 The Precision-Recall curves computed by LSH, SpH, ITQ, SDH, SGH, DPSH, ADSH, DAPH, and RADH on the IAPR TC-12 dataset. Figures from (**a**) to (**f**) are associated with the code length of 8-bit, 12-bit, 16-bit, 24-bit, 36-bit, and 48-bit 236

Fig. 7.13 The Precision-Recall curves computed by LSH, SpH, ITQ, SDH, SGH, DPSH, ADSH, DAPH, and RADH on the MIRFLICKR-25K dataset. Figures from (**a**) to (**f**) are associated with the code length of 8-bit, 12-bit, 16-bit, 24-bit, 36-bit, and 48-bit 238

Fig. 7.14 The sketch of the network for learning the single-image and cross-image representations 241

Fig. 7.15 The proposed deep architecture of the pairwise comparison model (best viewed in color) 246

Fig. 7.16 The proposed deep architecture of the triplet comparison model (best viewed in color) 247

Fig. 7.17 Rank-1 accuracy versus λ in the CUHK03 dataset (best viewed in color) 250

Fig. 7.18 The rank-1 accuracies and CMC curves of different methods on the CUHK03 dataset [37] (best viewed in color) 251
Fig. 7.19 The rank-1 accuracies and CMC curves of different methods on the CUHK01 dataset [37] (best viewed in color) 252
Fig. 7.20 The rank-1 accuracies and CMC curves of different methods on the VIPeR dataset [37] (best viewed in color) 253

List of Tables

Table 2.1 The average accuracy and error bar (percentage) of 5 independent experiments for DM detection 23
Table 2.2 The area under curve (AUC) for the different approaches in DM detection. Bold values mean the best performances 24
Table 2.3 The average accuracy and error bar (percentage) in 5 independent experiments for IGR detection. Bold values mean the best performances 27
Table 2.4 The area under curve (AUC) for the different methods in IGR detection. Bold values mean the best performances 28
Table 2.5 Face recognition rates on the Extended Yale B database. Bold values mean the best performances 32
Table 2.6 Face recognition rates on the AR database. Bold values mean the best performances 32
Table 2.7 Recognition accuracy by competing approaches on the disguise AR database. Bold values mean the best performances 33
Table 2.8 Face recognition rates on FRGC2.0 Exp 4. Bold values mean the best performances 34
Table 2.9 Face recognition rates on LFW. Bold values mean the best performances 34
Table 2.10 The categorization rates on the 17 class Oxford Flowers dataset. The results in bracket are obtained under equal feature weights. Bold values mean the best performances 36
Table 2.11 The categorization accuracy on the 102 category Oxford Flowers dataset. The results in bracket are obtained under equal feature weights. Bold values mean the best performances 36
Table 2.12 The classification rate and their corresponding error bar in 5 independent experiments for Fatty Liver detection. Best results are highlighted in bold 45

Table 2.13 The AUC values on the Fatty Liver diagnosis when the number of training instances is 200, 250, and 300, respectively. Best results are highlighted in bold 48
Table 2.14 The time consumption using the proposed method with different training data numbers 48
Table 3.1 The overall and mean classifying accuracy values calculated by SAGP and other comparison methods with the dimensionality of the latent variable varying from 1 to 10 using dataset Wiki Test-Image. Bold values mean the best performances 64
Table 3.2 The classifying accuracy values of each category use different individual or multi-view strategies on Wiki Test-Image dataset 65
Table 3.3 The overall and mean classifying accuracy values calculated by SAGP and other comparison approaches with the dimensionality of latent variable varying from 40 to 130 using dataset AWA. Bold values mean the best performances 66
Table 3.4 The overall and mean classifying accuracy values calculated by SAGP and other comparison methods with dimensionality of the latent variable varying from 1 to 30 using the dataset NUS-WIDE-LITE. Bold values mean the best performances 68
Table 3.5 The overall and average classification accuracies on the AWA dataset obtained by MKSGP and other comparison methods when the dimensionality of the latent variable changes from 40 to 130. Bold values mean the best performances 78
Table 3.6 The overall and average classifying accuracy values calculated by various approaches with q varying from 1 to 30 using dataset NUS-WIDE-LITE. Bold values mean the best performances 80
Table 3.7 The accuracy, sensitivity, and specificity values calculated by different methods on Biomedical dataset with the training samples numbers set 30, 40, and 50, respectively. Best results are highlighted with bold 82
Table 3.8 Overall/average classifying accuracy values calculated by SLEMKGP and various comparison strategies with dimensionality q varying from 1 to 10 using dataset Wiki Text-Image. Bold values mean the best performances 91
Table 3.9 Overall/average classifying accuracy values calculated by SLEMKGP and various comparison approaches with dimensionality q varying from 40 to 130 using AWA dataset 93
Table 3.10 Overall/average classifying accuracy values calculated by SLEMKGP and various comparison approaches with

dimensionality q varying from 1 to 30 using dataset NUS-WIDE-LITE 95
Table 3.11 Overall/average classifying accuracy values calculated by SLEMKGP with varying dimensionality d_p of $\{\mathbf{P}^v\}_{v=1}^V$ by using the datasets AWA and NUS-WIDE-LITE 97
Table 3.12 Overall/average classifying accuracy values calculated by SLEMKGP when $d_p = q$ using the datasets AWA and NUS-WIDE-LITE 97
Table 4.1 Some necessary notations in this chapter 104
Table 4.2 The classification accuracy on synthetic dataset with a different number of views and features. Bold values mean the best performances 113
Table 4.3 The classification accuracies on the synthetic dataset obtained by MVMFL. Bold values mean the best performances 115
Table 4.4 The average classification accuracies and error bars on the biomedical dataset obtained by MVMFL and other comparison methods when the number of training samples changes from 30 to 90. Bold values mean the best performances 116
Table 4.5 The area under curve (AUC) for the different methods in DM detection. Bold values mean the best performances 117
Table 4.6 The classification accuracies on the synthetic dataset obtained by HMMF. Bold values mean the best performances ... 127
Table 4.7 The accuracy, sensitivity, and specificity values obtained by different methods on the Biomedical dataset when the number of training samples is 40, 50, 60, and 70, respectively. Best results are highlighted in bold 128
Table 4.8 The area under curve (AUC) obtained by the different methods in DM detection. Bold values mean the best performances 129
Table 5.1 Description of 8 UCI datasets 146
Table 5.2 Comparisons with baselines of single metric and combined metrics with single feature on LFW dataset 149
Table 5.3 Comparisons with baselines of single metric and combined metrics with multiple features on LFW dataset 150
Table 5.4 Comparisons with the state-of-the-art metric learning methods on LFW dataset 151
Table 5.5 Accuracy (%) comparison with the state-of-the-art results on LFW dataset under image restricted protocol 153
Table 5.6 Accuracy (%) comparisons with existing deep metric learning on LFW data in restricted protocol 154
Table 5.7 Performance comparisons between our GMSL and other metric learning on Pubfig faces 157

Table 5.8 Descriptions of the 11 UCI and the 4 handwritten digit datasets 167
Table 5.9 Comparison of classification error rate between CDSM and CDSM with distance part (CDSM_{dis}), CDSM with similarity part (CDSM_{sim}). Here, μ is a balance parameter to control the effect of distance or similarity part in CDSM. Bold values mean the best performances 168
Table 5.10 Comparison of classification error rate between our CDSM and the state-of-the-art metric learning methods on the UCI datasets. Bold values mean the best performances 168
Table 5.11 Comparison of classification error rate between our CDSM and the state-of-the-art metric learning methods on the handwritten digit datasets. Bold values mean the best performances 169
Table 5.12 Comparison of classification error rate between CDSM and kernelized CDSM. Bold values mean the best performances 170
Table 5.13 Influence on running time and correctness by two improvement strategies. In counting the training time, we leave out the time for constructing triplets 171
Table 5.14 Running time of different metric learning methods on UCI datasets. Note that we compute the time for training stage without building constraints 171
Table 6.1 Classification error rate (%) of either of NIR and VIS images from the HFB dataset 181
Table 6.2 Classification error rate (%) of the proposed fusion approach and the conventional weighted matching score level fusion approaches on the HFB dataset 182
Table 6.3 Classification error rate (%) of either of the 2D and 3D palmprint images 183
Table 6.4 Classification error rates (%) of the proposed fusion approach and the conventional weighted matching score level fusion approaches on the 2D plus 3D palmprint dataset 183
Table 6.5 Classification error rate (%) of blue, green, or near infrared images of the Multispectral dataset 184
Table 6.6 Classification error rates (%) of the proposed fusion approach and the conventional weighted matching score level fusion approaches on the Multispectral dataset with the score fusion of blue and near infrared images 184
Table 6.7 Classification error rates (%) of the proposed fusion approach and the conventional weighted matching score level fusion approaches on the Multispectral dataset with the score fusion of green and near infrared images 184

Table 6.8 Classification error rate (%) of the R, G, B color channels of the GT face dataset 185
Table 6.9 Classification error rates (%) of the proposed fusion approach and the conventional weighted matching score level fusion approaches on the GT face dataset based on the score fusion of the red and green channels 186
Table 6.10 Classification error rates (%) of the proposed fusion approach and the conventional weighted matching score level fusion approaches on the GT face dataset based on the score fusion of the blue and green channels 186
Table 6.11 Classification error rates (%) of the proposed fusion approach and the conventional weighted matching score level fusion approaches on the GT face dataset based on the score fusion of the red and blue channels 186
Table 6.12 Classification error rate (%) of the R, G, B color channels of the LFW face dataset 187
Table 6.13 Classification error rates (%) of the proposed fusion approach and the conventional weighted matching score level fusion approaches on the LFW face dataset based on the score fusion of the red and green channels. 187
Table 6.14 Classification error rates (%) of the proposed fusion approach and the conventional weighted matching score level fusion approaches on the LFW face dataset based on the score fusion of the blue and green channels 187
Table 6.15 Classification error rates (%) of the proposed fusion approach and the conventional weighted matching score level fusion approaches on the LFW face dataset based on the score fusion of the red and blue channels 188
Table 6.16 Summary of the fourteen datasets 192
Table 6.17 Classification accuracy (%) of Lib-SVM, FaLK-SVM, FaLK-SVMad, and FaLKSVMar on the fourteen UCI datasets 193
Table 6.18 Summary of the three large-scale datasets 193
Table 6.19 Classification accuracy (%) of Lib-SVM, FaLK-SVM, FaLK-SVMad, and FaLKSVMar on the three large-scale datasets 194
Table 7.1 The network structure of CNN-F 201
Table 7.2 The MAP scores obtained by different methods on the IAPR TC-12 dataset. Bold values mean the best performances 208
Table 7.3 The Top-500 MAP and Top-500 Precision scores obtained by different methods on the IAPR TC-12 dataset. Bold values mean the best performances 209

Table 7.4 The MAP scores obtained by different methods on the MIRFLICKR-25K dataset. Bold values mean the best performances 211
Table 7.5 The Top-500 MAP and Top-500 Precision scores obtained by different methods on the MIRFLICKR-25K dataset. Bold values mean the best performances 212
Table 7.6 The MAP scores obtained by different methods on the CIFAR-10 dataset. Bold values mean the best performances 214
Table 7.7 The Top-500 MAP and Top-500 Precision scores obtained by different methods on the CIFAR-10 dataset. Bold values mean the best performances 215
Table 7.8 The MAP scores obtained by different methods on the CIFAR-10 dataset, respectively. 1000 samples are selected for testing and remaining samples are used for training in this dataset. Bold values mean the best performances 217
Table 7.9 The MAP scores obtained by different methods on the CIFAR-10 dataset. Bold values mean the best performances 230
Table 7.10 The MAP@Top500 and Precision@Top500 scores obtained by different methods on the CIFAR-10 dataset. Bold values mean the best performances 230
Table 7.11 The MAP scores obtained by different methods on the NUS-WIDE dataset. Bold values mean the best performances 232
Table 7.12 The MAP@Top500 and Precision@Top500 scores obtained by different methods on the NUS-WIDE dataset. Bold values mean the best performances 232
Table 7.13 The MAP scores obtained by different methods on the IAPR TC-12 dataset. Bold values mean the best performances 234
Table 7.14 The MAP@Top500 and Precision@Top500 scores obtained by different methods on the IAPR TC-12 dataset. Bold values mean the best performances 234
Table 7.15 The MAP scores obtained by different methods on the MIRFLICKR-25K dataset. Bold values mean the best performances 237
Table 7.16 The MAP@Top500 and Precision@Top500 scores obtained by different methods on the MIRFLICKR-25K dataset. Bold values mean the best performances 237
Table 7.17 The rank-1 accuracies (%) of the proposed pairwise and triplet comparison models 250
Table 7.18 The training times of the proposed pairwise and triplet comparison models 250