SPARSE TIKHONOV-REGULARIZED HASHING FOR MULTI-MODAL LEARNING

Lei Tian*,†,‡, *Xiaopeng Hong**,‡, *Chunxiao Fan*†, *Yue Ming*†, *Matti Pietikäinen**, *Guoying Zhao*$^{\star,\diamond}$

* The Center for Machine Vision and Signal Analysis, University of Oulu, Finland, 90014
† Beijing University of Posts and Telecommunications, Beijing, P.R.China, 100876
‡ Co-First author $^{\diamond}$ Corresponding author

ABSTRACT

This paper mainly focuses on the role of regularization in Multi-Modal Learning (MML). Existing MML studies devote most of the efforts in maximizing the consensus of models from cues of different modalities. However, regularization methods are still far from fully explored. To fill in this gap, we propose a compact and efficient coding solution, termed by Sparse Tikhonov-Regularized Hashing (STRH). The STRH enforces both the ℓ_0-norm induced sparsity constraints and the Tikhonov regularization on the binary solution vectors which maximize cross-modal correlation. In addition, we raise the concerns on the challenging testing scenario of 'Multi-modal Learning and Single-modal Prediction' (MLSP). Finally, we demonstrate that the STRH is an efficient hashing solutions by showing its superiority under the MLSP scenario.

Index Terms— Multi-Modal Learning, Binary Hashing, Tikhonov Regularization, ℓ_0-norm Sparsity Constraint

1. INTRODUCTION

Multi-Modal Learning (MML) has received rapidly growing attention in image and video acquisitions and processing techniques. Existing MML studies devote most of the efforts in maximizing the consensus from cues of different modalities. However, regularization methods are still far from fully explored, in spite that they have been shown as effective ways of introducing additional information to solve particular problems such as ill-posed or overfitting. Therefore, we mainly focus on the role of regularization in multi-modal learning and present a compact and efficient coding solution, termed by Sparse Tikhonov-Regularized Hashing (STRH).

It is worth mentioning that traditional MML studies usually investigate in an ideal 'Multi-modal Learning and Multi-modal Prediction' scenario (MLMP) [1, 2, 3, 4, 5]. However, in practice, very often the modalities used in the learning phase are partly absent in prediction phase. To allow a more practical study of MML, we raise the concerns on the testing scenario of 'Multi-modal Learning and Single-modal Prediction' (MLSP).

The proposed method is clearly connected to the study of MML. According to different formats of output, existing MML methods can be categorized into two types: **real-valued** and **hashing MML**.

Real-valued MML: Two representative *unsupervised* real-valued MML methods are Canonical Correlation Analysis (C-CA) [6] and Partial Least Squares (PLS) [7]. *Supervised* real-valued MML methods incorporate the label information to obtain a set of discriminative projection matrices, such as the Generalized Multi-view Analysis (GMA) [1], and Multi-view Discriminant analysis (MvDA) [2].

Hashing MML: The hashing MML methods output binary codes for further recognition and analysis tasks [5, 8, 9, 3]. Just as two typical examples, the work [5] combines the intra-modality loss terms and ℓ_2 regularization. Wu *et al* [3] consider both hashing function learning and hashing quantization.

Most multi-modal learning researches mainly focus on the cross-modality connection [8, 4, 10, 11]. In contrast, we formulate the objective function by treating 1) cross-modality connection, 2) ℓ_0-norm induced sparsity, 3) Tikhonov-regularization, and 4) binary constraints, to facilitate the exploration of regularization in MML.

The contributions of this paper include: 1) We propose STRH multi-modal learning method. Its objective function unifies cross-modality connection term and within-modality regularization terms, which consists of the ℓ_0-norm induced sparsity, Tikhonov-regularization, and binary constraints. 2) As solving the strict ℓ_0 regularization is NP-hard, we derive a computationally feasible solution for this complex optimization problem. 3) We demonstrate that the STRH is an efficient hashing solutions with good computationally stability and group sparsity under the MLSP scenario.

2. SPARSE TIKHONOV-REGULARIZED HASHING

Let $\mathbb{Z}$ denote a set of N training samples, each of which is associated with T modalities and *one* label. Its i-th sample is represented by $Z_i = \{\mathbf{x}_{i,1}, \cdots, \mathbf{x}_{i,t}, \cdots, \mathbf{x}_{i,T}, l_i\}$, where $i = 1, \cdots, N$, $l_i \in L \subset \mathbb{R}$ is a label, and $\mathbf{x}_{i,t} \in \chi^t \subset \mathbb{R}^{d_t}$ is a d_t-dimensional modality-specific feature vector extracted from the t-th modality for $t = 1, \cdots, T$.

Our goal is to learn a set of mapping functions $\Phi = \{\phi_1, \cdots, \phi_t, \cdots, \phi_T\}$, where $\phi_t : \chi^t \to \mathbb{R}^K$, to minimize the cross-modality inconsistency $\mathfrak{L}(\Phi, \mathbb{Z})$ under a group of regularization constraints $\mathfrak{R}(\phi_t(\cdot))$. $\mathfrak{R}(\phi_t(\cdot))$ are defined on the modality-specific mappings. We introduce two auxiliary matrices $\mathbf{X}_t = [(\mathbf{x}_{1,t}), \cdots, (\mathbf{x}_{N,t})] \in \mathbb{R}^{d_t \times N}$ and $\mathbf{\Psi}_t = [\phi_t(\mathbf{x}_1), \cdots, \phi_t(\mathbf{x}_N)] \in \mathbb{R}^{K \times N}$, the columns of which are the output vectors of ϕ when applied to N samples. K is the size of binary codes.

Cross-modal Connection: The loss function $\mathfrak{L}$ is designed to model the semantic cross-modality connection. We simplify $\mathfrak{L}$ as the modality-pair-wise loss function:

$$\mathfrak{L}(\Phi, \mathbb{Z}) = -a_{i,j}\phi_s^{\mathrm{T}}(\mathbf{x}_i)\,\phi_r(\mathbf{x}_j), \tag{1}$$

where $a_{i,j}$ is a label agreement indicator, i.e., $a_{i,j} = 1$ when $l_i = l_j$, otherwise $a_{i,j} = -1$. In the matrix form, the minimizing of $\mathfrak{L}(\Phi, \mathbb{Z})$ becomes a maximizing problem:

$$\max_{\Phi} \sum_{s,r} \mathrm{tr}\left(\mathbf{\Psi}_s \mathbf{A} \mathbf{\Psi}_r^T\right), \tag{2}$$

where tr$(\cdot)$ is the matrix trace. $\mathbf{A} \in \mathbb{R}^{N \times N}$ is a label matrix and it emphasizes the consistency of multiple modalities. E-q. 2 is maximized when the output of different modalities are highly correlated for those samples with same labels.

Regularization: Then we construct the robust regularization terms to introduce extra enhancement in computational stability, sparsity, and compactness.

Encoding high-dimensional features using binary hashes of low dimension is useful for fast similarity computations, with few performance sacrifice [12, 13, 14]. We define the modality mapping functions as:

$$\phi_t(\mathbf{x}_i) = \mathrm{sgn}\left(\mathbf{V}_t^T \mathbf{x}_i\right), \tag{3}$$

where $\mathbf{V}_t \in \mathbb{R}^{d_t \times K}$ is the linear projection matrix for the t-th modality data, and sgn$(\cdot)$ is the signum function.

The Tikhonov-regularization (i.e., ℓ_2-regularization) has shown its ability in computational stability for ill-posed problems [15, 16]. Moreover, enforcing a sparsity constraint leads to simpler and more interpretable solutions [17, 13]. Compared with the ℓ_1-norm induced sparsity, the ℓ_0-norm constraint not only provides a more sensible sparsity constraint but also enables to control the sparsity straightforwardly. Thus, the ℓ_0-induced sparsity and the ℓ_2-regularizations are leveraged as our sparsity and stability constraints.

2.1. Objective Function

There are two terms in our model, *i.e*, the cross-modality connection term (Eq. 2) and within-modality regularizations including ℓ_0-sparsity, the Tikhonov-regularization, and binary constraints (Eq. 3). We formulate objective function as:

$$\max_{\Phi} \sum_{s,r} \mathrm{tr}\left(\mathrm{sgn}\left(\mathbf{V}_s^T \mathbf{X}_s\right) \cdot \mathbf{A} \cdot \mathrm{sgn}\left(\mathbf{X}_r^T \mathbf{V}_r\right)\right), \tag{4}$$
$$\text{s.t. } \left|\mathbf{V}_t\right|_0 \le m_t, \ \ \left\|\mathbf{V}_t\right\|^2 \le \varepsilon_t, \ \ t = 1, \cdots, T,$$

where $\Phi = \{\mathbf{V}_1, \mathbf{V}_2 \cdots, \mathbf{V}_T\}$, $|\cdot|_0$ is the number of non-zero elements of a matrix, and $\|\cdot\|^2$ is the Frobenius-norm, which induces the Tikhonov-regularization. We simply set $m_1 = \cdots = m_T = (1-p) \times (d_t K)$, where p is the sparsity.

For clarity, hereinafter we take the *two-modality* case as an example to optimize Eq. 4[1].

2.2. Optimization

The sgn function and the ℓ_0-norm make Eq. 4 NP-hard to solve. We rely on the iterative *variable splitting and penalty* optimization techniques [17, 13] to find a feasible solution. Given $\mathbf{X}$ and $\mathbf{Y}$ as data matrices of two modalities, their projection matrices are denoted by $\mathbf{V}$ and $\mathbf{W}$, respectively. Eq. 4 can be specified as follow:

$$\max_{\mathbf{V},\mathbf{W}} \ \mathrm{tr}\left(\mathrm{sgn}\left(\mathbf{V}^T\mathbf{X}\right) \cdot \mathbf{A} \cdot \mathrm{sgn}\big(\mathbf{Y}^T\mathbf{W}\big)\right) \tag{5}$$
$$\text{s.t. } |\mathbf{V}|_0 \le m_1, \ \|\mathbf{V}\|^2 \le \varepsilon_1, \ |\mathbf{W}|_0 \le m_2, \ \|\mathbf{W}\|^2 \le \varepsilon_2.$$

Considering the symmetric roles of $\mathbf{V}$ and $\mathbf{W}$ in Eq. 5, we split the optimization into two subproblems namely **Problem-V** and **Problem-W** by solving one single-modal projection matrix and keeping the other one fixed iteratively:

$$\begin{aligned} &\max_{\mathbf{V},\mathbf{W}} \ \ \mathrm{tr}\left(\mathrm{sgn}\left(\mathbf{V}^T\mathbf{X}\right) \cdot \mathbf{A} \cdot \mathbf{C}_V\right), \\ &\text{s.t.} \ \ \ |\mathbf{V}|_0 \le m_1, \ \|\mathbf{V}\|^2 \le \varepsilon_1, \end{aligned} \tag{6}$$

and

$$\begin{aligned} &\max_{\mathbf{V},\mathbf{W}} \ \ \mathrm{tr}\left(\mathbf{C}_W \cdot \mathbf{A} \cdot \mathrm{sgn}\big(\mathbf{W}^{\mathbf{T}}\mathbf{Y}\big)^T\right), \\ &\text{s.t.} \ \ \ |\mathbf{W}|_0 \le m_2, \ \|\mathbf{W}\|^2 \le \varepsilon_2, \end{aligned} \tag{7}$$

where $\mathbf{C}_V = \mathrm{sgn}\big(\mathbf{Y}^T\mathbf{W}\big)$ and $\mathbf{C}_W = \mathrm{sgn}\left(\mathbf{V}^T\mathbf{X}\right)$ are fixed for each of the subproblems. Obviously it is similar to optimize these two objective functions. For clarity, we only describe the solution to **Problem-V**.

Problem-V is still highly challenging because of the *discrete binary constraint*. Inspired by [18], an binary auxiliary matrix $\mathbf{B} \in \{-1, 1\}^{K \times N}$ is introduced to separate the optimization of the projection matrix and discrete binary constraint. Eq. 6 then becomes

$$\begin{aligned} &\max_{\mathbf{B}_X,\mathbf{V}} \ \ \mathrm{tr}\left(\mathbf{B}_X \cdot \mathbf{A} \cdot \mathbf{C}_V\right) - \lambda_1\left\|\mathbf{B}_X - \mathbf{V}^T X\right\|^2 \\ &\text{s.t. } |\mathbf{V}|_0 \le m_1, \ \|\mathbf{V}\|^2 \le \varepsilon_1, \ \mathbf{B}_X \in \{-1, 1\}^{K \times N} \end{aligned} \tag{8}$$

where λ_1 is a penalty parameter. Thus, two alternative steps are obtained for optimizing Eq. 8: updating the discrete binary codes $\mathbf{B}_X$ and updating the sparse projection matrix $\mathbf{V}$.

2.2.1. Updating $\mathbf{B}_X$

By fixing $\mathbf{V}$, Eq. 8 can be expanded and easily obtain an optimal analytical solution of $\mathbf{B}_X$ as:

$$\mathbf{B}_X = \mathrm{sgn}\left(\mathbf{C}_V^T \cdot \mathbf{A}^T + 2\lambda_1 \mathbf{V}^T \mathbf{X}\right). \tag{9}$$

[1] It can be easily extended to the more general *multiple-modality* case by updating one variable and fixing the others on a rota basis.

2.2.2. *Updating* $\mathbf{V}$

With $\mathbf{B}_X$ updated, it is still difficult to solve $\mathbf{V}$ by the ℓ_0-induced sparsity and the ℓ_2 constraints jointly in Eq. 8. We apply the *variable splitting and penalty* techniques again, and split the projection matrix $\mathbf{V}$ into two same-size matrices $\mathbf{V}_1$ and $\mathbf{V}_2$, constrained by the ℓ_0-norm and ℓ_2-norm constraints. Given $\mathbf{B}_X$ fixed, Eq.8 is expressed as follows:

$$\begin{aligned}\min_{\mathbf{V}_1,\mathbf{V}_2} \quad & \lambda_1\left\|\mathbf{B}_X - \mathbf{V}_2^T\mathbf{X}\right\|^2 + \alpha_1\left\|\mathbf{V}_1^T\mathbf{X} - \mathbf{V}_2^T\mathbf{X}\right\|^2 + \beta_1\|\mathbf{V}_2\|^2 \\ \text{s.t.} \quad & |\mathbf{V}_1|_0 \leq m_1\end{aligned} \tag{10}$$

where α_1 is sparsity term related penalty. Note that we convert the ℓ_2-norm constraint into the objective function and β_1 is a corresponding parameter. In the following paragraphs, we solve Eq. 10 in an alternate manner.

Updating sparsity regularization $\mathbf{V}_1$: Let $\mathbf{V}_2$ be fixed. Similar to [13], we firstly ignore the sparsity constraint and expand the objective function as

$$J = \text{tr}\left(\mathbf{V}_1^T\mathbf{X}\mathbf{X}^T\mathbf{V}_1 - 2\mathbf{V}_2^T\mathbf{X}\mathbf{X}^T\mathbf{V}_1 + \mathbf{V}_2^T\mathbf{X}\mathbf{X}^T\mathbf{V}_2\right). \tag{11}$$

We iteratively optimize it by gradient descent. The sparsity constraint can be satisfied by thresholding the gradient descent based solution and keeping the m_1 largest elements. The rest elements are set to be 0. Therefore,

$$\mathbf{V}_1^{(\tau+1)} = \text{th}_{m_1}\left(\mathbf{V}_1^{(\tau)} - \gamma\mathbf{X}\mathbf{X}^T\left(\mathbf{V}_2 - \mathbf{V}_1^{(\tau)}\right)\right), \tag{12}$$

where $\text{th}_{m_1}(\mathbf{Z})$ keeps the largest m_1 elements of matrix $\mathbf{Z}$ and γ denotes the learning rate.

Updating Tikhonov-regularization $\mathbf{V}_2$: Let $\mathbf{V}_1$ be fixed. Benefit from the ℓ_2-norm term, Eq. 10 has a closed-form solution by setting the derivative of its expansion to 0.

$$\mathbf{V}_2 = \left[(\lambda_1 + \alpha_1)\mathbf{X}\mathbf{X}^T + \beta_1\mathbf{I}\right]^{-1}\left(\lambda_1\mathbf{X}\mathbf{B}_X^T + \alpha_1\mathbf{X}\mathbf{X}^T\mathbf{V}_1\right), \tag{13}$$

where $\mathbf{I}$ denotes identity matrix of size $d_1 \times d_1$.

We can also solve **Problem-W** by analogy[2].

3. EXPERIMENTS

We evaluate the proposed STRH in this section, and we choose six representative methods, i.e., CCA [6] and PLS [7] for unsupervised real-valued MML, GMA (GMA-MFA) [1] and MvDA [2] for supervised real-valued MML, SCM [4] and QCH [3] for hashing-based MML.

We test STRH using different combinations of modalities on two datasets. The Pascal VOC 2007 *image retrieval* dataset [19] consists of 2,954 training and 3,192 testing image-tag pairs with single label. The SMIC *micro-expression (ME)* dataset [20] contains 71 clips with three heterogeneous images, i.e., high-speed (HS), normal (VIS) and near-infrared (NIR) images. It is obviously a dataset of small size samples (SSS), and thus provides a good opportunity to demonstrate that the introduction of the *group sparsity* term can solve the over-fitting problem.

Table 1. Comparison results (%) for HMER task under MLSP scenario. Scenario 1 and 2 denote ***(VIS+HS) learning & VIS prediction*** and ***(VIS+NIR) learning & VIS prediction*** .

	CCA	**PLS**	**GMA**	**MvDA**	**SCM**	**QCH**	**Ours**
Scenario 1	69.01	61.97	54.93	61.97	63.38	53.52	**73.66**
Scenario 2	69.01	59.15	54.93	61.97	56.34	52.11	**73.24**

3.1. Implementation Details

Pascal dataset: We evaluate MML methods under ***(image + tag) learning & image prediction*** MLSP scenario. The visual feature extraction for image retrieval usually contains two steps [21]: 1) feature extraction from original images and 2) feature encoding. For the feature extraction, we extract a set of 766d hand-crafted features [21] and a 4096d deep learning feature from the fc7 layer of AlexNet [22] model for each image. For feature encoding, we choose three methods, i.e., LSH [23], ITQ [24] and DPSH [25]. In order to show our STRH does not depend on the text-like descriptors either, we test three tag features: Absolute Tag Rank (ATR), Relative Tag Rank (RTR), and Word Frequency (WF).

For the proposed STRH method, we set the parameters as $p = 0.5$, $\lambda_1 = \lambda_2 = 100$, $\beta_1 = \beta_2 = 1$, $\alpha_1 = \alpha_2 = 10$ according to the cross-validation. We evaluate the performance for dimensionality of output $K = \{16, 32, 64, 128, 256, 512\}$, and report the results in terms of mean of Average Precision (mAP).

SMIC dataset: We use the HIGO-TOP feature [26, 27] for Heterogeneous Micro-Expression Recognition (HMER) task [27]. We test two modality combinations namely 1) ***(VIS + HS) learning & VIS prediction***, 2) ***(VIS + NIR) learning & VIS prediction***, to investigate how STRH improves the performance of VIS modality by utilizing extra NIR and HS modalities. For the SSS issue, we increase the sparsity p to 0.7 and use the same parameters with the above experiments.

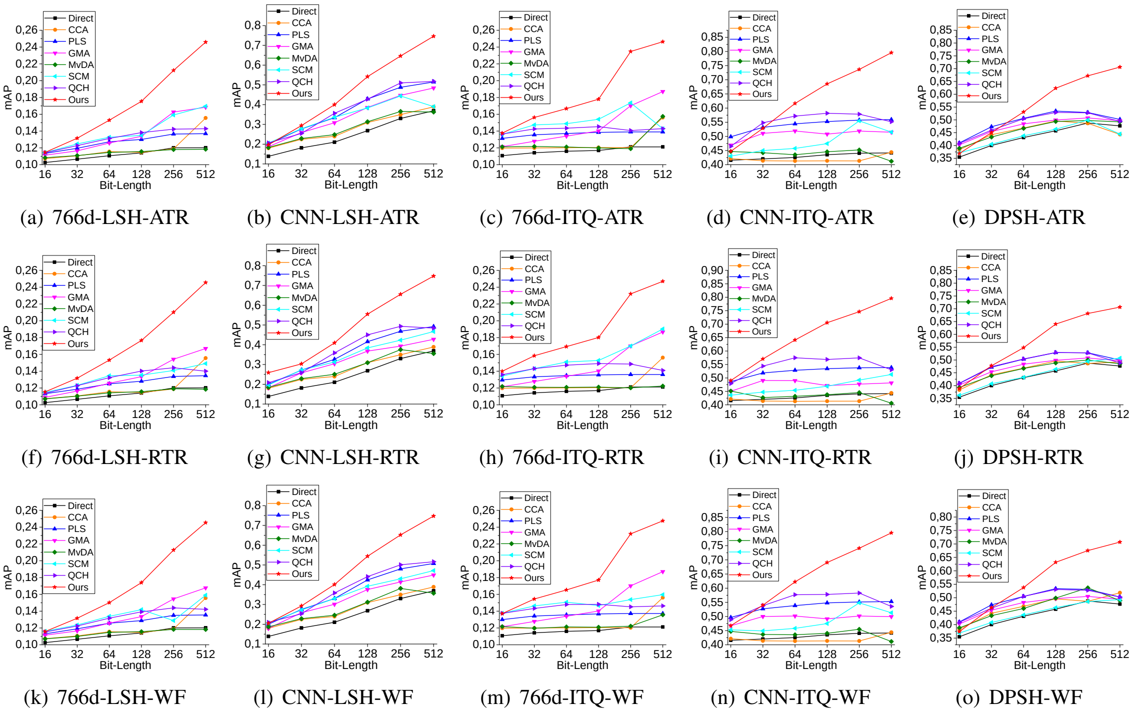
3.2. Experimental Results

Pascal dataset: Fig. 1 shows the mAP scores of the six MML methods, using 15 different features as input. The STRH clearly dominates the comparison under various experimental settings over almost all values of K. The performance edge becomes more evident on larger K.

SMIC dataset: The results of HMER are listed in Table 1. It can be easily observed that the STRH substantially outperforms other methods under both *Scenario 1* and *2*.

From the reported results, it is safe to reach the following conclusions: 1) the increase in accuracy brought by STRH is independent of the input feature types, no matter on which

[2] The detailed derivation is similar to the **Problem-V** on a rota basis



(a) 766d-LSH-ATR (b) CNN-LSH-ATR (c) 766d-ITQ-ATR (d) CNN-ITQ-ATR (e) DPSH-ATR

(f) 766d-LSH-RTR (g) CNN-LSH-RTR (h) 766d-ITQ-RTR (i) CNN-ITQ-RTR (j) DPSH-RTR

(k) 766d-LSH-WF (l) CNN-LSH-WF (m) 766d-ITQ-WF (n) CNN-ITQ-WF (o) DPSH-WF

Fig. 1. Comparison results in term of **mAP** on Pascal dataset. These experiments are performed under ***(image + tag) learning & image prediction*** MLSP scenario.

datasets it is tested; 2) The introduction of regularization makes STRH capture the cross-modality connection in a more effective and stable manner. 3) By using the sparsity constraint, STRH effectively reduces the risk of over-fitting and thus provides an efficient solution to the SSS problem.

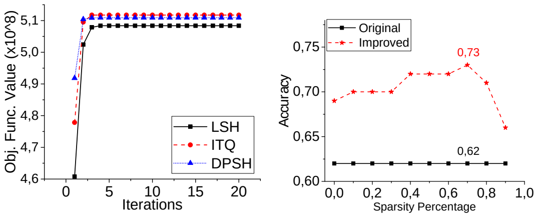


Fig. 2. **Left**: Convergence of the objective function on Pascal dataset. **Right**: The impact of sparsity on SMIC dataset.

3.3. Discussion

We investigate two key issues, i.e., convergence and sparsity. **Convergence:** We use the hand-crafted feature on Pascal dataset and keep track of the changes of the objective function values along iterations. As shown by the left part of Fig. 2, it shows that STRH basically converges within five iterations, regardless of input features.
Sparsity Percentage: To further demonstrate the necessity of sparsity term for the SSS dataset, we evaluate the sparsity percentage under the *scenario 1* on the SMIC dataset. The right part of Fig.2 shows that the accuracy moderately increases together with the sparsity percentage. It demonstrates that the introduction of sparsity constraint indeed improves the robustness.More importantly, our model provides a mechanism to balance the over-fitting and under-fitting risk by controlling the sparsity percentage p.

4. CONCLUSIONS

In this paper, we propose a sparse Tikhonov-regularized hashing method for multi-modal learning. It enforces both the ℓ_0-norm induced sparsity constraints and the Tikhonov regularization on the binary solution vectors which maximize cross-modal correlation. We provide a computationally feasible solution to STRH and investigate it under the challenging MLSP test scenario that we suggest. The STRH outperforms popular MML methods, regardless of the types of input features.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Grants No. 61402046, 61572205, 61772419), the Academy of Finland, Infotech Oulu, Tekes Fidipro program (Grant No. 1849/31/2015) and Tekes project (Grant No. 3116/31/2017).

5. REFERENCES

[1] Abhishek Sharma, Abhishek Kumar, Hal Daume, and David W Jacobs, "Generalized multiview analysis: A discriminative latent space," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2160–2167.

[2] Meina Kan, Shiguang Shan, Haihong Zhang, Shihong Lao, and Xilin Chen, "Multi-view discriminant analysis," *IEEE TPAMI*, vol. 38, no. 1, pp. 188–194, 2016.

[3] Botong Wu, Qiang Yang, Wei-Shi Zheng, Yizhou Wang, and Jingdong Wang, "Quantized correlation hashing for fast cross-modal search.," in *IJCAI*, 2015, pp. 3946–3952.

[4] Dongqing Zhang and Wu-Jun Li, "Large-scale supervised multimodal hashing with semantic correlation maximization.," in *AAAI*, 2014, vol. 1, p. 7.

[5] Yi Zhen and Dit-Yan Yeung, "Co-regularized hashing for multimodal data," in *NIPS*, 2012, pp. 1376–1384.

[6] John Shawe-Taylor and Nello Cristianini, *Kernel methods for pattern analysis*, Cambridge university press, 2004.

[7] Herman Wold, "Partial least squares," *Encyclopedia of statistical sciences*, 1985.

[8] Deming Zhai, Hong Chang, Yi Zhen, Xianming Liu, Xilin Chen, and Wen Gao, "Parametric local multimodal hashing for cross-view similarity search.," in *IJCAI*, 2013, pp. 2754–2760.

[9] Xianglong Liu, Junfeng He, Cheng Deng, and Bo Lang, "Collaborative hashing," in *CVPR*, 2014, pp. 2139–2146.

[10] Di Wang, Xinbo Gao, Xiumei Wang, and Lihuo He, "Semantic topic multimodal hashing for cross-media retrieval.," in *IJCAI*, 2015, pp. 3890–3896.

[11] Kun Ding, Bin Fan, Chunlei Huo, Shiming Xiang, and Chunhong Pan, "Cross-modal hashing via rank-order preserving," *IEEE TMM*, vol. 19, no. 3, pp. 571–585, 2017.

[12] Fumin Shen, Chunhua Shen, Wei Liu, and Heng Tao Shen, "Supervised discrete hashing," in *CVPR*, 2015, pp. 37–45.

[13] Yan Xia, Kaiming He, Pushmeet Kohli, and Jian Sun, "Sparse projections for high-dimensional binary codes," in *CVPR*, 2015, pp. 3332–3339.

[14] Wang-Cheng Kang, Wu-Jun Li, and Zhi-Hua Zhou, "Column sampling based discrete supervised hashing.," in *AAAI*, 2016, pp. 1230–1236.

[15] Ahmed Alaoui and Michael W Mahoney, "Fast randomized kernel ridge regression with statistical guarantees," in *NIPS*, 2015, pp. 775–783.

[16] Yichao Lu, Paramveer Dhillon, Dean P Foster, and Lyle Ungar, "Faster ridge regression via the subsampled randomized hadamard transform," in *NIPS*, 2013, pp. 369–377.

[17] Yilun Wang, Junfeng Yang, Wotao Yin, and Yin Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM Journal on Imaging Sciences*, vol. 1, no. 3, pp. 248–272, 2008.

[18] Fumin Shen, Wei Liu, Shaoting Zhang, Yang Yang, and Hengtao Shen, "Learning binary codes for maximum inner product search," in *ICCV*, 2015, pp. 4148–4156.

[19] Mark Everingham, Luc Van Gool, Christopher K-I Williams, John Winn, and Andrew Zisserman, "The pascal visual object classes (voc) challenge," *IJCV*, vol. 88, no. 2, pp. 303–338, 2010.

[20] Xiaobai Li, Tomas Pfister, Xiaohua Huang, Guoying Zhao, and Matti Pietikäinen, "A spontaneous micro-expression database: Inducement, collection and baseline," in *FG*. IEEE, 2013, pp. 1–6.

[21] Go Irie, Hiroyuki Arai, and Yukinobu Taniguchi, "Alternating co-quantization for cross-modal hashing," in *ICCV*, 2015, pp. 1886–1894.

[22] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *NIPS*, 2012, pp. 1097–1105.

[23] Alexandr Andoni and Piotr Indyk, "Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions," in *FOCS*. IEEE, 2006, pp. 459–468.

[24] Yunchao Gong, Svetlana Lazebnik, Albert Gordo, and Florent Perronnin, "Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval," *IEEE TPAMI*, vol. 35, no. 12, pp. 2916–2929, 2013.

[25] Wu-Jun Li, Sheng Wang, and Wang-Cheng Kang, "Feature learning based deep supervised hashing with pairwise labels," *arXiv preprint arXiv:1511.03855*, 2015.

[26] Xiaopeng Hong, Yingyue Xu, and Guoying Zhao, "Lbp-top: a tensor unfolding revisit," in *ACCV*. Springer, 2016, pp. 513–527.

[27] Xiaobai Li, HONG Xiaopeng, Antti Moilanen, Xiaohua Huang, Tomas Pfister, Guoying Zhao, and Matti Pietikainen, "Towards reading hidden emotions: A comparative study of spontaneous micro-expression spotting and recognition methods," *IEEE TAC*, 2017.