

Hessian Regularized Support Vector Machines for Mobile Image Annotation on the Cloud

Dapeng Tao, Lianwen Jin, *Member, IEEE*, Weifeng Liu, *Member, IEEE*, and Xuelong Li, *Fellow, IEEE*

Abstract—With the rapid development of the cloud computing and mobile service, users expect a better experience through multimedia computing, such as automatic or semi-automatic personal image and video organization and intelligent user interface. These functions heavily depend on the success of image understanding, and thus large-scale image annotation has received intensive attention in recent years. The collaboration between mobile and cloud opens a new avenue for image annotation, because the heavy computation can be transferred to the cloud for immediately responding user actions. In this paper, we present a scheme for image annotation on the cloud, which transmits mobile images compressed by Hamming compressed sensing to the cloud and conducts semantic annotation through a novel Hessian regularized support vector machine on the cloud. We carefully explained the rationality of Hessian regularization for encoding the local geometry of the compact support of the marginal distribution and proved that Hessian regularized support vector machine in the reproducing kernel Hilbert space is equivalent to conduct Hessian regularized support vector machine in the space spanned by the principal components of the kernel principal component analysis. We conducted experiments on the PASCAL VOC'07 dataset and demonstrated the effectiveness of Hessian regularized support vector machine for large-scale image annotation.

Index Terms—Cloud computing, Hessian Eigenmaps and support vector machines, manifold regularization, mobile service.

I. INTRODUCTION

TODAY, smart phones equipped with a digital camera have become more and more popular, and personal digital photo is easily produced in massive quantities. Although

Manuscript received March 02, 2012; revised June 25, 2012; accepted November 13, 2012. Date of publication January 11, 2013; date of current version May 13, 2013. This work was supported by the National Basic Research Program of China (973 Program) under Grant 2012CB316400, the National Natural Science Foundation of China under Grants 61075021, 61271407, 61125106, and 91120302, the Guangdong Natural Science Funds under Grant S2011020000541, the Guangdong Scientific and Technology Research Plan under Grants 2012A010701001 and 2011B090400146, the Fundamental Research Funds for the Central Universities of China, South China University of Technology, under Grant 2012ZP0002, and the Natural Science Foundation for Youths of Shandong Province, China, under Grant ZR2011FQ016. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Joel Rodrigues.

D. Tao and L. Jin are with the School of Electronic and Information Engineering, South China University of Technology, GuangZhou 510640, Guangdong, China (e-mail: dapeng.tao@gmail.com; lianwen.jin@gmail.com).

W. Liu is with the College of Information and Control Engineering, China University of Petroleum, China (e-mail: liuwf@upc.edu.cn).

X. Li is with the Center for OPTical IMagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shaanxi, China (e-mail: xuelong_li@opt.ac.cn).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMM.2013.2238909

it is popular to combine time and directory for the photo management in smart phones, it is inconvenient to effectively retrieve photos at the semantic level. Therefore, a large number of image annotation based systems [12], [14], [27], [33], [35] have been developed by utilizing the semantic keywords for the personal photo organization.

Image annotation aims to assign several key words to an image. It is one of the most fundamental research problems in image processing, computer vision and multimedia [19]. Typically, it is accomplished by the following procedure. Given a collection of training images, we first extract visual features to represent these images [17], [41]. Afterward, a set of models are trained based on these images for the subsequent annotation, each of which corresponds to a particular key word (or concept).

Recently, various approaches based on clustering algorithms [43], support vector machines (SVM) [3], [37] [55], [63], manifold learning algorithms [31], [42] [56] have been proposed to resolve the image annotation task. Labeled samples are effective to improve the annotation performance. However, it is difficult to get a large number of labeled samples. Thus semi-supervised learning is promising to improve the annotation performance. Motivated by recent progress in Hessian Eigenmaps [16] and SVM, in this paper, we introduce the Hessian regularization to improve SVM for large-scale image annotation.

Given limited computational resource in a mobile, in general, it is impossible to annotate a newly captured image in an online fashion. In addition, the small storage in a mobile does not allow users to maintain a large number of pictures to form a reliable manifold regularization. Therefore, it is essential to consider an alternative way to annotate a mobile image.

With the rapid development of the Internet technology, cloud computing [38], [58] becomes mature and already offers dozens of services [26], [38] [52], [54] to the community, in general including 1) Cloud Infrastructures as a Service (IaaS) that provides the users with storage, networks and other high scalable computing resource, e.g., virtual computer servers, 2) Cloud Platform as a Service (PaaS) that provides APIs to the development of applications on an abstract platform, and 3) Cloud Software as a Service (SaaS) that provides applications running on a cloud infrastructure to various client devices through a thin client interface, e.g., a web browser [38]. Cloud computing is rapidly changing the landscape of information technology [1], [60], as well as attracting more and more academic attentions such as service-oriented science [22], [36], online image processing services [50], [51], handwriting recognition services [23], and storage service for storing scientific data [8]. Since scientific computing usually requires a large number of resources

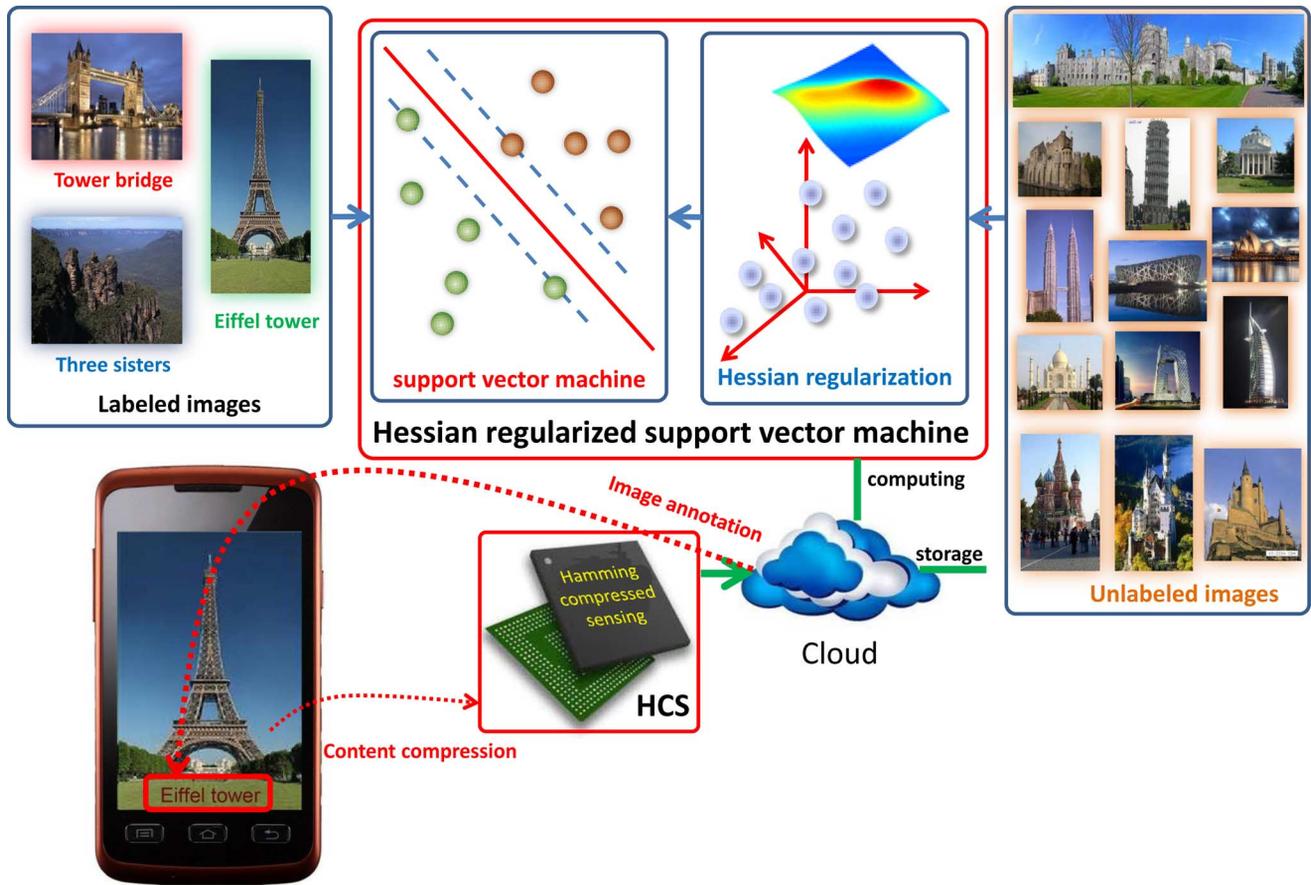


Fig. 1. Image annotation on the cloud. This scheme contains the following five components: 1) using Hamming compressed sensing (HCS) to compress a mobile image and transmit the compressed image to the cloud, 2) using HCS to decode the compressed image in the cloud, 3) training the Hessian regularized support vector machines by using a small number of labeled images and a large number of unlabeled images stored in the cloud, 4) annotating the mobile image by using the trained Hessian regularized support vector machines and 5) return the annotation results to the mobile terminal.

to deliver results for ever-growing problem size in an acceptable time, the cloud computing technology provides significant benefit for scientific computing community by offering cheap alternative to supercomputers, a much more reliable platform than grids, and a much more scalable platform than clusters [36]. In addition, it deals with the trouble raised by client terminals diversity of Mobile Internet comfortably. Thus, it is natural to transmit the mobile image to the cloud for annotation and storage. By utilizing cloud computing, we can easily handle image annotation requests of mobile devices with different operation systems, such as iOS, Android, Windows Phone, and Blackberry. Thus, image annotation can be provided to the end user as a kind of SaaS cloud service. Especially, the proposed Hessian regularized support vector machines which involves large computing resources are ready applicable in the cloud IaaS platform and can be easily implemented in a parallel fashion to fully utilize the strength of the cloud computing.

The success of compressed sensing [9], [15] opens an innovative channel to effectively and efficiently compress mobile images, so we can save cost for transmitting mobile images to the cloud. Note that in contrast to conventional compressed sensing and 1-bit compressed sensing [7], Hamming compressed sensing (HCS) [62] is more suitable for the proposed application, because HCS can directly recover a digital signal (the quantized real-valued signal) from the quantization

of its linear measurements. HCS is based on the observation of the statistics of 1 and -1 in the 1-bit measurements from the viewpoint of computational geometry. The effectiveness of HCS is ensured by the strong theoretical guarantee for the small recovery error bound and the small number of measurements required by successful recovery.

Based on the above descriptions, precisely annotating images captured by a mobile camera or stored in a mobile become reality through the following stages: 1) training the Hessian regularized support vector machines by using a small number of labeled images and a large number of unlabeled images stored in the cloud, 2) using HCS to compress a mobile image and transmit the compressed image to the cloud, 3) using HCS to decode the compressed image in the cloud, 4) annotating the mobile image by the trained Hessian regularized support vector machines and 5) return the annotation results to the mobile.

Fig. 1 shows the architecture of the proposed mobile image annotation on the cloud. The most main client terminal is a smartphone with Android or iOS installed. The communication channel between client terminals and the cloud supports the popular socket protocol, such as WiFi, EDGE, 3G and 4G networks. The bandwidth of popular wireless communication protocols is basic suit for the data of image compressed by Hamming compressed sensing (HCS). In general, less of time delay of transmission will make customers satisfied. It feels like that

the annotation is done locally. The main computing process of the image annotation are actually all provisioned in the cloud, including HCS decoding, training the Hessian regularized SVM with both a small number of labeled images and a large number of unlabeled images stored in the cloud, and annotating the image using the trained Hessian regularized SVM. With the benefit of the rapid elasticity capability of cloud computing, the computing and storage resource can be easily and quickly scale out to meet the increasing requirement when there is a rise in image annotation service demand.

The main contribution of this paper is the newly developed Hessian regularized support vector machine for large-scale image annotation. Given the limited page length, the other parts, which are easy to implement based on the related references, will not be detailed in this paper.

The rest of the paper is organized as follows: In Section II, we present related works on image annotation. In Section III, we detail the newly proposed Hessian regularized support vector machines. Section IV shows the experimental results on the PASCAL VOC challenge 2007. Section V concludes the paper.

II. RELATED WORK

In recent years, a dozen of effective methods have been proposed to automatically annotate images, which can be grouped into three categories according to the used machine learning schemes, including unsupervised, supervised and semi-supervised learning.

Unsupervised learning algorithms, such as topic model [2], [39], nonnegative matrix factorization [29] and clustering [43], have been introduced to annotate images. For example, Barnard *et al.* [2] presented a scheme of probabilistic latent variable model to infer relations between visual features of images and the associated texts. Extensive numerical experiments [39] verified the appealing accuracy, consistency and efficiency of latent space models. In contrast to latent variable models that need to model sample distribution properly, similarity based clustering algorithms do not assume the distribution of the image set. Thus normalized cut (NCut) [48] and its extensions become popular in practice [11], [18]. Recently, Papadopoulos *et al.* [43] presented a clustering scheme for the detection of landmarks and events.

Supervised learning, such as support vector machines (SVM) [57], parallel least squares [24] and decision trees [44], based image annotation explores the relationship between tags and image visual features via a set of training samples. For example, Wang *et al.* [59] presented a novel supervise topic model [6] to predict class labels and annotation terms simultaneously. Shi *et al.* [49] proposed the text-based Bayesian model (TBM) to improve Bayesian Hierarchical Multinomial Mixture Models (BH-MMMs), since the original annotations are incomplete to estimate the accurate model parameter. Since visual features are distributed on a low dimensional manifold embedded in a high dimensional ambient space, many manifold learning based dimension reduction tools [31], [35] have been introduced to annotate images. For example, Lin *et al.* [35] constructed the relational graphs through human computer interactions to encode similar and dissimilar image pairs. Maximum margin projec-

tion (MMP) [31] finds both the local geometry of a within-class graph and discriminant information of a between-class graph. Although active learning [28] can be introduced to interactively select only effective unlabeled samples for labeling, the user labeling effort is heavy. In general, supervised learning algorithms require a large number of manually labeled samples to obtain robust annotation performance.

In practice, it is impossible to get a large number of labeled samples. Considering the availability of large amount of unlabeled samples, e.g., social images on Flickr, Picasa and Photobucket Pro, semi-supervised image annotation has received intensive attention. In particular, the unlabeled samples are used to explore the data distribution that can be deemed as a prior to control the model complexity. This strategy essentially enhances the image annotation performance. Bilenko *et al.* [5] developed a new framework that learns distance-function learning by exploring constrains. This method is a variant of k-means clustering. Shao *et al.* [47] presented a semi-supervised algorithm that can be treated as a combination of statistical inference and manifold learning. Thus, in some way its performance is superior to the methods belonging to mere latent space models.

Although existing semi-supervised learning algorithms [13], [64] have improved image annotation a lot, they are mainly based upon the well-known Laplacian regularization [3], [4]. Recently, it has been observed that the Laplacian regularization biases the solution towards a constant function [34], [52]. In contrast to the Laplacian regularization, we introduce the Hessian regularization to image annotation, which drives the learned function varying linearly along the data manifold. In particular, we treat Hessian Eigenmaps [16] as a manifold regularization for encoding the local geometry of unlabeled samples to improve support vector machines for large-scale image annotation. That is because Hessian Eigenmaps calculate the tangent Hessian to approximate the Hessian energy in a computationally effective way.

III. HESSIAN REGULARIZED SUPPORT VECTOR MACHINES

Support vector machines (SVM) [57] are an innovative and effective classification method developed under the theme of the statistical learning theory [57]. It has attractive advantages in solving small sample learning, data nonlinearity and data high dimensionality problems, and thus received intensive attentions in recent years.

A dozen variants of SVM have been developed to handle different problems, such as transductive SVM, parallel SVM, Nesterov SVM [63], multi-view SVM, Laplacian regularized SVM [4], ensemble manifold regularized SVM [25] and active SVM [55].

In this paper, we present a new variant of SVM, Hessian regularized SVM. It uses a large number of unlabeled samples to construct the Hessian regularization to boost the classification performance of SVM, when the number of labeled samples is small. Although it is similar to Laplacian regularized SVM [4], it significantly improves Laplacian regularized SVM. That is because Hessian regularization drives the learned function varying linearly along the data manifold.

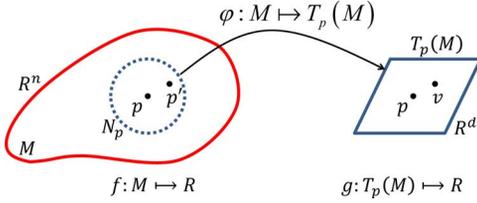


Fig. 2. The tangent Hessian of f at the point p in $T_p(M)$ is defined as the ordinary Hessian of $g: T_p(M) \mapsto R$, because the smooth mapping φ uniquely maps $p' \in N_p$ on M to $v \in T_p(M)$. The coordinate system of $T_p(M)$ at $p \in M$ can be estimated by the eigenspace of the neighborhood N_p of p associated with the largest d eigenvalues.

Given a small number of labeled images $L = \{(x_i, y_i)\}_{i=1}^l$ with $y_i \in \{+1, -1\}$ and a large number of unlabeled images $U = \{x_j\}_{j=l+1}^{l+u}$, Hessian regularized SVM learns a classification function to separate positive samples from negative samples. Under the theme of semi-supervised learning, it is general to assume labeled images $(x, y) \in \mathbb{R}^n \times \{\pm 1\}$ are drawn from P and unlabeled images $x \in \mathbb{R}^n$ are drawn from the marginal distribution P_X of P . This marginal distribution P_X can help the function learning if P_X is related to the conditional distribution $P(y|x)$. The underlying assumption here is close image pairs x_i and x_j mean similar conditional distribution pairs $P(y|x_i)$ and $P(y|x_j)$. The manifold assumption is widely used in computer vision and multimedia applications. It assumes close image pairs induce similar conditional distribution pairs. It is ideal to adopt this assumption on images containing single object. In real-world applications, each image usually contains multiple objects, but the manifold assumption still helps to improve the performance. Therefore, it is important to discover the intrinsic geometry of the manifold $M \subset \mathbb{R}^n$, i.e., the support of P_X , from which samples are drawn.

In this Section, we first introduce Hessian regularization induced by the unlabeled images, which is derived from Hessian Eigenmaps [16]. We then present Hessian regularized SVM. Afterward, we show rationality of Hessian regularization by comparing with graph Laplacian. Finally we show Hessian regularized SVM in RKHS is equal to conducting Hessian regularized linear SVM in the space spanned by the principal components of kernel principal component analysis.

A. Hessian Regularization

Given a smooth manifold $M \subset \mathbb{R}^n$, we can define the tangent space $T_p(M) \subset \mathbb{R}^n$ for point $p \in M$. In order to obtain the Hessian of a function $f: M \mapsto R$, we need to define the orthonormal coordinate system of $T_p(M)$ which can be approximated by the eigenspace spanned by the neighborhood N_p of p associated with the largest d eigenvalues (this d essentially determines the intrinsic dimension of the manifold M and it can be selected in an empirical way, e.g., according to the distribution of the eigenvalues). For each point $p' \in N_p$, we can find a unique closest point v on $T_p(M)$, i.e., there is a smooth mapping φ that uniquely maps $p' \in N_p$ on M to $v \in T_p(M)$. Therefore, the Hessian of f at the point p in $T_p(M)$ can be defined as the ordinary Hessian of $g: T_p(M) \mapsto R$. Thus, Donoho and Crimes [16] termed it as tangent Hessian. Fig. 2 illustrates

the above process to define the tangent Hessian, mathematically defined by

$$(H_f^{\text{tan}}(p))_{i,j} = \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} g(x)|_{x=0}. \quad (1)$$

The construction of the Hessian matrix of a point depends on the choice of the coordinate system in $T_p(M)$. Since M formed by the large number of unlabeled images is not linear, the coordinate system in $T_p(M)$ varies along the manifold. Fortunately, the Frobenius norm of a Hessian matrix is invariant to the coordinate changes, and therefore we can have the Hessian regularization to encode the average curviness of f along the manifold M , i.e.,

$$H(f) = \int_{p \in M} \|H_f^{\text{tan}}(p)\|_F^2 dp. \quad (2)$$

It has been shown by Donoho and Crimes in [16] that $H(f)$ has a $d+1$ dimensional nullspace spanned by the constant function and the eigenspace of the neighborhood N_p of p associated with the largest d eigenvalues. Thus, we brief the discretization of the Hessian regularization for encoding the local geometry of unlabeled samples in the following steps [34]:

- Step 1: Finding the k -nearest neighbors N_p of the j -th unlabeled sample x_j and centralizing the neighborhood by taking x_j off from the k -nearest neighbors. This centralization makes x_j to be the origin of the tangent space $T_{x_j}(M)$. This is different from Hessian Eigenmaps [16] that use the average of all the k -nearest neighbors and x_j as the origin. The advantage of this method is that the j -th unlabeled sample x_j is exactly a sample drawn from the intrinsic manifold. In contrast, Donoho's method uses the mean of x_j and its k -nearest neighbors as the center that makes the local tangent space biased.
- Step 2: Estimating the orthonormal coordinate system of the tangent space $T_{x_j}(M)$ by the eigenspace U of the neighborhood N_p of p associated with the largest d eigenvalues. This can be implemented by conducting singular value decomposition on $X^j = [x_i - x_j]_{i=1}^k$, where in x_i is the i -th sample in the k -nearest neighbors N_p .
- Step 3: Taking the $(d+1)$ -dimensional nullspace off from the matrix $H^j = [1, u_1 \dots u_m, u_1 u_1 \dots u_d u_d]$ by using Gram-Schmidt orthonormalization, and resulting \bar{H}^j . The Frobenius norm of \bar{H}^j is given by $(\bar{H}^j)^T \bar{H}^j$.
- Step 4: Accumulating $(\bar{H}^j)^T \bar{H}^j$ over all images and then resulting the Hessian regularization $f^T H f$.

B. Hessian Regularized Support Vector Machines

The same as the ordinary SVM that takes hinge loss, Hessian regularized SVM for binary classification thus takes the form

$$\begin{aligned} f^* &= \arg \min_{f \in H_K} \frac{1}{l} \sum_{i=1}^l L(x_i, y_i) + \lambda_K \|f\|_K^2 + \lambda_H \|f\|_H^2 \\ &= \arg \min_{f \in H_K} \frac{1}{l} \sum_{i=1}^l (1 - y_i f(x_i))_+ + \lambda_K \|f\|_K^2 \\ &\quad + \lambda_H \|f\|_H^2, \end{aligned} \quad (3)$$

where $L(x_i, y_i)$ is the loss function and takes the form $(1 - y_i f(x_i)) = \max(0, 1 - y_i f(x_i))$ for obtaining the margin maximization, $\|f\|_k^2$ is the reproducing kernel Hilbert space (RKHS) H_k norm of f , λ_k controls the complexity of f in the ambient space R^n , $\|f\|_H^2$ is the Hessian regularization to encode the intrinsic geometry of marginal distribution P_X , and λ_H controls the complexity of f in compact support of P_X .

Theorem 1: The minimization of Hessian regularized SVM w.r.t. $f \in H_K$ exists and takes the following representation

$$f(x) = \sum_{i=1}^n \alpha_i^* K(x_i, x), \quad (4)$$

where $n = l + u$.

This representer theorem indicates that f can be completely characterized by an expansion of both labeled and unlabeled samples. It shows the existence and the general form of the solution of Hessian regularized SVM. The proof of this representer theorem is almost standard in kernel machines, e.g., representer theorem proved in [4], so we only sketch it in the following lines.

Proof: The space spanned by the kernels centered at labeled and unlabeled samples is $S = \text{span}\{K(x, \cdot) | x \in L \cup U\}$ and its orthogonal complement is S^\perp . Thus, any $f \in H_K$ can be written as $f = f_S + f_{S^\perp}$, wherein f_S is the projection of f to S and f_{S^\perp} is the projection of f to S^\perp .

The inner product associated with H_K is defined by $\langle f, g \rangle = \int_{L \cup U} f(x)g(x)dP_X$ and thus $\|f\|_H^2 = \langle f, Hf \rangle$. For any f and g , if $f - g|_M = 0$, then we have $\|f\|_H^2 = \|g\|_H^2$.

Since S^\perp is the orthogonal complement of S and S is spanned by kernels centered at labeled and unlabeled samples drawn from M , any functions in S^\perp vanishes on M . This results in for $\forall x \in M$, $f(x) = f_S(x) + f_{S^\perp}(x) = f_S(x)$, i.e., $\|f\|_H^2 = \|f_S\|_H^2$. Therefore, $f(x) = \sum_{i=1}^n \alpha_i^* K(x_i, x)$. ■

From (3), by introducing the slack variables ς_i for $1 \leq i \leq l$, we have

$$\begin{aligned} f^* = \arg \min_{\alpha \in R^n, \varsigma \in R^l} & \frac{1}{l} \sum_{i=1}^l \varsigma_i + \lambda_K \alpha^T K \alpha + \frac{\lambda_H}{n^2} \alpha^T K H K \alpha \\ \text{s.t. } & y_i \left(\sum_{j=1}^n \alpha_j^* K(x_j, x_i) + b \right) \geq 1 - \varsigma_i, \quad 1 \leq i \leq l \\ & \varsigma_i \geq 0, \quad 1 \leq i \leq l. \end{aligned} \quad (5)$$

Similar to Laplacian SVM [3], by using the Lagrangian method, the optimal α^* is given by

$$\alpha = \left(2\lambda_K I + \frac{2\lambda_H}{n^2} H K \right)^{-1} [I \ 0]^T Y \beta^*, \quad (6)$$

where I is the $l \times l$ identity matrix, 0 is the $l \times u$ zero matrix, $Y = \text{diag}(y_1, y_2, \dots, y_l)$ and β^* is given by

$$\begin{aligned} \beta^* = \arg \min_{\beta \in R^l} & \sum_{i=1}^l \beta_i - \frac{1}{2} \beta^T Q \beta \\ \text{s.t. } & \sum_{i=1}^l \beta_i y_i = 0 \\ & 1 \leq \beta_i \leq \frac{1}{l}, \quad 1 \leq i \leq l \end{aligned} \quad (7)$$

where $Q = Y[I \ 0]K(2\lambda_K I + 2\lambda_H/n^2 H K)[I \ 0]^T Y$ with I is the $l \times l$ identity matrix and 0 is the $l \times u$ zero matrix.

Problem (7) is a standard quadratic programming and can be solved by standard SVM solvers, such as LIBSVM [10], NeSVM [63], SVM-light [40], SVM-Perf [32], and parallel SVM. Different SVM solvers converge at different rates and have particular advantages. For example, NeSVM converges at rate $O(1/k^2)$, SVM-Perf converges at rate $O(1/\sqrt{k})$, and Pegasos [46] converges at rate $O(1/k)$. For mobile image annotation on the cloud, we can consider parallel SVM which distributes training data on parallel machines (in the cloud), and reduces memory requirement through parallel row-based Incomplete Cholesky Factorization (ICF) on the loaded data. Its time cost is about $O(np^2/m)$ and its space cost is about $O(np/m)$, wherein n is the problem size (the number of samples for constructing the kernel matrix), p is the reduced dimensionality obtained by ICF and m is number of machines in the cloud to conduct the parallel computing.

C. Rationality of Hessian Regularization

Fitting the hinge loss function for labeled images is an essential ill-posed problem under the theme of regularization theory, so we need regularizations $\|f\|_K^2$ and $\|f\|_H^2$ to smooth the fitting function to avoid the wild oscillation. Since an RKHS is a Hilbert space of functions of the form $f(x) = \sum_{i=1}^n \alpha_i K(x_i, x)$ shown in the representer theorem, $\|f\|_K^2$ essentially controls the capacity of RKHS. Thus, it becomes popular in kernel machines. In contrast to existing kernel classification models, $\|f\|_H^2$, originally used in dimension reduction [16] and linear regression (with different form according to Eells-energy [20] but share similar functional [34]), smoothes the classification function f along the manifold and further restricts the space of admissible solutions by considering the distribution of unlabeled samples (the underlying assumption is both the unlabeled and labeled samples are independent and identically distributed). The combination of $\|f\|_K^2$ and $\|f\|_H^2$ ensures the solution has good predictive capabilities for the subsequent classification.

In Hessian regularized SVM, since the marginal distribution P_X is unknown in practice, it is essential to empirically estimate a suitable penalty to encode the smoothness along the manifold supporting P_X . The Hessian regularization $\|f\|_H^2$ is such a suitable one because it encodes the local geometry of the compact support of P_X .

Although graph Laplacian $L(f) = \int_M \|\nabla_M f\|_F^2$ [4] has been introduced to classification and regression and has shown promising support for improving the performance, its nullspace is the constant function along the compact support of P_X . Since the function f will not be penalized in the nullspace, $L(f)$ drives f towards a constant function. In contrast to $L(f)$, the Frobenius norm of the tangent Hessian $H(f) = \int_M \|\nabla_M^2 f\|_F^2$ has the null space spanned by a constant function and a d -dimensional original isometric coordinates. This fact has been well explored by Eells and Lemaire [20], then by Donoho and Crimes [16], and recently by Steinke and Hein [52]. This property ensures the Hessian regularization drives the learned classification f varying linearly along the data manifold, because it penalizes the second derivative along the manifold and ignores the normal direction. In addition, it has been

shown [16] that Hessian Eigenmaps can handle non-convex situations, such as a manifold containing holes, comparing with ISOMap [53] and locally linear embedding (LLE) [45]. According to patch alignment for dimension reduction [61], we know popular manifold learning algorithms encode different kinds of local geometry which can be well characterized on the tangent coordinate. Roughly, Hessian Eigenmaps is different to other manifold learning algorithms, such as ISOMap, LLE, Laplacian Eigenmaps [3], and Local Tangent Space Alignment, because it is of second order while others are of first order. Therefore, the Hessian regularization $H(f) = \int_M \|\nabla_M^2 f\|_F^2$ is superior to the Laplacian regularization $L(f) = \int_M \|\nabla_M f\|_F^2$ for both classification and regression [34], and has the potential to explore the intrinsic local geometry of the compact support of P_X .

In contrast to the Eells-energy used to penalize least squares for linear regression [34], in which Eells-energy is a kind of Hessian regularization, we use the Frobenius norm of the tangent Hessian induced from Hessian Eigenmaps [16] for smoothing SVM in RKHS along the compact support of the conditional distribution P_X .

D. Understanding RKHS

This small section presents an understanding to the RKHS through the space spanned by the principal components of the kernel principal components (KPCA). In particular, we prove that Hessian regularized SVM in RKHS is equal to conducting Hessian regularized linear SVM in the space spanned by the principal components of KPCA.

Theorem 2: Hessian regularized support vector machines in RKHS are equivalent to conducting Hessian regularized linear support vector machines in the principal components of KPCA.

Proof: Given labeled and unlabeled samples X_l (built by samples in L) and X_u (built by samples in U), the covariance matrix in RKHS is

$$C = \frac{1}{n} \sum_{i=1}^{l+u} \phi(x_i, \cdot) \phi(x_j, \cdot)^T. \quad (8)$$

For KPCA, we have eigenvector \mathbf{u} and the corresponding eigenvalue λ satisfying $C\mathbf{u} = \lambda\mathbf{u}$. This is equal to solving

$$\phi(x_i, \cdot)^T C \mathbf{u} = \lambda \phi(x_i, \cdot)^T \mathbf{u}, \quad i = 1, \dots, l+u. \quad (9)$$

We can prove that $u = \sum_{i=1}^{l+u} \beta_i \phi(x_i, \cdot)$, which is similar to the proof of the representer theorem. Then (9) is equivalent to following optimization problem:

$$K\beta = \lambda\beta, \quad (10)$$

where $K_{ij} = k(x_i, x_j)$, the solution of (9) is the eigenvector $\beta_i = (\beta_{i,1}, \dots, \beta_{i,l+u})$ and the corresponding eigenvalue λ_i . The projection matrix of KPCA is given by $U = (\mathbf{u}, \dots, \mathbf{u}_{l+u})$, each

$$\mathbf{u}_i = \sum_{j=1}^{l+u} \beta_{i,j} \phi(x_j, \cdot) = \Phi \beta_i, \quad (11)$$

where $\Phi = [\phi(x_1, \cdot), \dots, \phi(x_{l+u}, \cdot)]$.

Normalize $\beta_i \leftarrow \beta_i / (\|\beta_i\| \sqrt{\lambda_i})$, then we have $\mathbf{u}_i^T \mathbf{u}_i = \beta_i^T \Phi^T \Phi \beta_i = \beta_i^T K \beta_i = \lambda_i \beta_i^T \beta_i = 1$ and $\mathbf{u}_i^T \mathbf{u}_j = 0$.

Thus, we have $U^T U = 1$.

Let $B = [\beta_1, \dots, \beta_{l+u}]$, since $U^T U = (\Phi B)^T \Phi B = B^T K B$, we obtain $B^T K B = 1$. This results in $B^T B = K^{-1}$. Consequently, the feature of projected data in the KPCA space (spanned by principal components of KPCA) is given by $\hat{x} = U^T \Phi(x_i, \cdot) = B^T \mathbf{k}$, where \mathbf{k} is the i -th column of K .

Therefore, the data preprocessed by KPCA is $\hat{X} = B^T K$.

The optimization problem of Hessian regularized linear SVM is

$$\min_w \frac{1}{l} \sum_{i=1}^l (1 - y_i f(x_i))_+ + \lambda_K \|w\|_2^2 + \frac{\lambda_H}{(u+l)^2} w^T X_{lu} H X_{lu}^T w. \quad (12)$$

The predict function is

$$f(x) = w^T x.$$

Ordinarily, a bias term b is added to the above form.

The primal problem:

$$\begin{aligned} \min_w \frac{1}{l} \sum_{i=1}^l \xi_i + \lambda_K \|w\|_2^2 + \frac{\lambda_H}{(u+l)^2} w^T X_{lu} H X_{lu}^T w \\ \text{s.t. } y_i (w^T x_i + b) \geq 1 - \xi_i, \quad i = 1, \dots, l \\ \xi_i \geq 0, \quad i = 1, \dots, l. \end{aligned} \quad (13)$$

Introducing the Lagrange multipliers μ_i, η_i , we have

$$\begin{aligned} L(w, \xi, b, \mu, \eta) \\ = \frac{1}{l} \sum_{i=1}^l \xi_i - \sum_{i=1}^l \eta_i \xi_i \\ + \frac{1}{2} w^T \left(2\lambda_K I + \frac{2\lambda_H}{(l+u)^2} X_{lu} H X_{lu}^T \right) w \\ - \sum_{i=1}^l \mu_i (y_i (w^T x_i + b) - 1 + \xi_i). \end{aligned} \quad (14)$$

Take the derivatives with respect to b and ξ_i , we obtain

$$\begin{aligned} \sum_{i=1}^l \mu_i y_i = 0 \frac{1}{l} - \mu_i - \eta_i = 0 \Rightarrow \mu_i \\ = \frac{1}{l} - \eta_i \quad (\eta_i \text{ is non-negative}) \Rightarrow 0 \\ \leq \mu_i \leq \frac{1}{l}. \end{aligned}$$

Using these identities, we formulate a reduced Lagrangian:

$$\begin{aligned} L^R(w, \mu) = \frac{1}{2} w^T \left(2\lambda_K I + \frac{2\lambda_H}{(l+u)^2} X_{lu} H X_{lu}^T \right) w \\ - \sum_{i=1}^l \mu_i (y_i w^T x_i - 1) \\ = \frac{1}{2} w^T \left(2\lambda_K I + \frac{2\lambda_H}{(l+u)^2} X_{lu} H X_{lu}^T \right) w \\ - w^T X_{lu} J^T Y \mu + \sum_{i=1}^l \mu_i \end{aligned}$$

where $j = [I \ 0]$ with I is the $l \times l$ identity matrix and 0 is the $l \times u$ zero matrix, and $Y = \text{diag}(y_1, \dots, y_l)$.

Taking the derivative of the reduced Lagrangian with respect to w :

$$\frac{\partial L^R}{\partial w} = \left(2\lambda_K I + \frac{2\lambda_H}{(l+u)^2} X_{lu} H X_{lu}^T \right) w - X_{lu} J^T Y \mu, \quad (15)$$

where H is the Hessian matrix, so it is symmetric.

This implies:

$$w = \left(2\lambda_K I + \frac{2\lambda_H}{(l+u)^2} X_{lu} H X_{lu}^T \right)^{-1} X_{lu} J^T Y \mu^*. \quad (16)$$

Substituting back in the reduced Lagrangian we get

$$\begin{aligned} \mu^* &= \max_{\mu} \sum_{i=1}^l \mu_i - \frac{1}{2} \mu^T Q \mu \\ \text{s.t.} \quad &\sum_{i=1}^l \mu_i y_i = 0 \\ &0 \leq \mu_i \leq \frac{1}{l}, \quad i = 1, \dots, l \end{aligned} \quad (17)$$

where

$$Q = Y J X_{lu}^T \left(2\lambda_K I + \frac{2\lambda_H}{(l+u)^2} X_{lu} H X_{lu}^T \right)^{-1} X_{lu} J^T Y.$$

Since the data preprocessed by KPCA is $\hat{X} = B^T K$, then we have the equation at the bottom of the page. The predict function is

$$\begin{aligned} f(\hat{x}) &= w^T \hat{x} = \hat{x}^T w \\ &= \mathbf{k}^T B \left(2\lambda_A I + \frac{2\lambda_H}{(l+u)^2} X_{lu} H X_{lu}^T \right)^{-1} \\ &\quad \times X_{lu} J^T Y \mu^* \\ &= \mathbf{k}^T B \left(2\lambda_A I + \frac{2\lambda_H}{(l+u)^2} B^T K H K B \right)^{-1} \\ &\quad \times B^T K J^T Y \mu^* \\ &= \mathbf{k}^T \tilde{\alpha}. \end{aligned} \quad (18)$$

This completes the proof. \blacksquare

IV. EXPERIMENTAL RESULTS

We evaluate the proposed Hessian regularized SVM on the PASCAL VOC'07 [21], which have been popularly used to evaluate image annotation and object recognition algorithms [30]. In this paper, we use the GIST feature extracted by Guillaumin *et al.* [30]. The performance is measured by using the average precision (AP) criterion for each class and the mean AP (mAP) over all classes. Details of the experimental setup and baseline models are given below.

A. the Setting of the Cloud Computing

We ran a set of experiments on a private cloud computing platform built on the OpenStack system. OpenStack is an open-source project that provides tools to build and manage public and private clouds. We used ten high performance computer workstations and high speed switch to build the private cloud computing system. Each workstation was configured with 6 Cores Xeon 2.4 GHz CPUs, 64G RAM, 2T hard disk and 1 Gbps networks interface card. Two of the workstations run the OpenStack Object Storage Service, one workstation runs the OpenStack Identity Service and Image Manager Service, one workstation runs the OpenStack Web Service, and the rest six workstations run the OpenStack Compute and Instance Services. The system installed Windows Server 2008 can launch 18 high performance virtual computer instances (4 Core VCPU, 16 G RAM, 600 G Hard disk), or at least 72 ordinary virtual computer instances (1 Core VCPU, 4 G RAM, 100 G Hard disk), in less than 30 minutes. The cloud computing experimental environment can satisfy the parallel computing experiments, handling the load balancing issue, support flexible use of the computing resources including virtual servers and storage. In our training stage, we use only five high performance virtual computers and 600 G cloud storage to contact all the experiments.

B. PASCAL VOC'07

In our image annotation experiments, we use the PASCAL VOC'07 dataset [21] that contains social images collected from the Flickr website. This widely used PASCAL VOC'07 dataset includes 9,963 images which were approximately equally divided into training and test subsets. The images in

$$\begin{aligned} \hat{Q} &= Y J \hat{X}_{lu}^T \left(2\lambda_K I + \frac{2\lambda_H}{(l+u)^2} \hat{X}_{lu} H \hat{X}_{lu}^T \right)^{-1} \hat{X}_{lu} J^T Y \\ &= Y J K B \left(2\lambda_K I + \frac{2\lambda_H}{(l+u)^2} B^T K H K B \right)^{-1} B^T K J^T Y \\ &= Y J K \left(2\lambda_K (B B^T K)^{-1} + \frac{2\lambda_H}{(l+u)^2} (B^T K)^{-1} B^T K H K B B^{-1} \right)^{-1} J^T Y \\ &= Y J K \left(2\lambda_K I + \frac{2\lambda_H}{(l+u)^2} H K \right)^{-1} J^T Y \\ &= \tilde{Q}. \end{aligned}$$

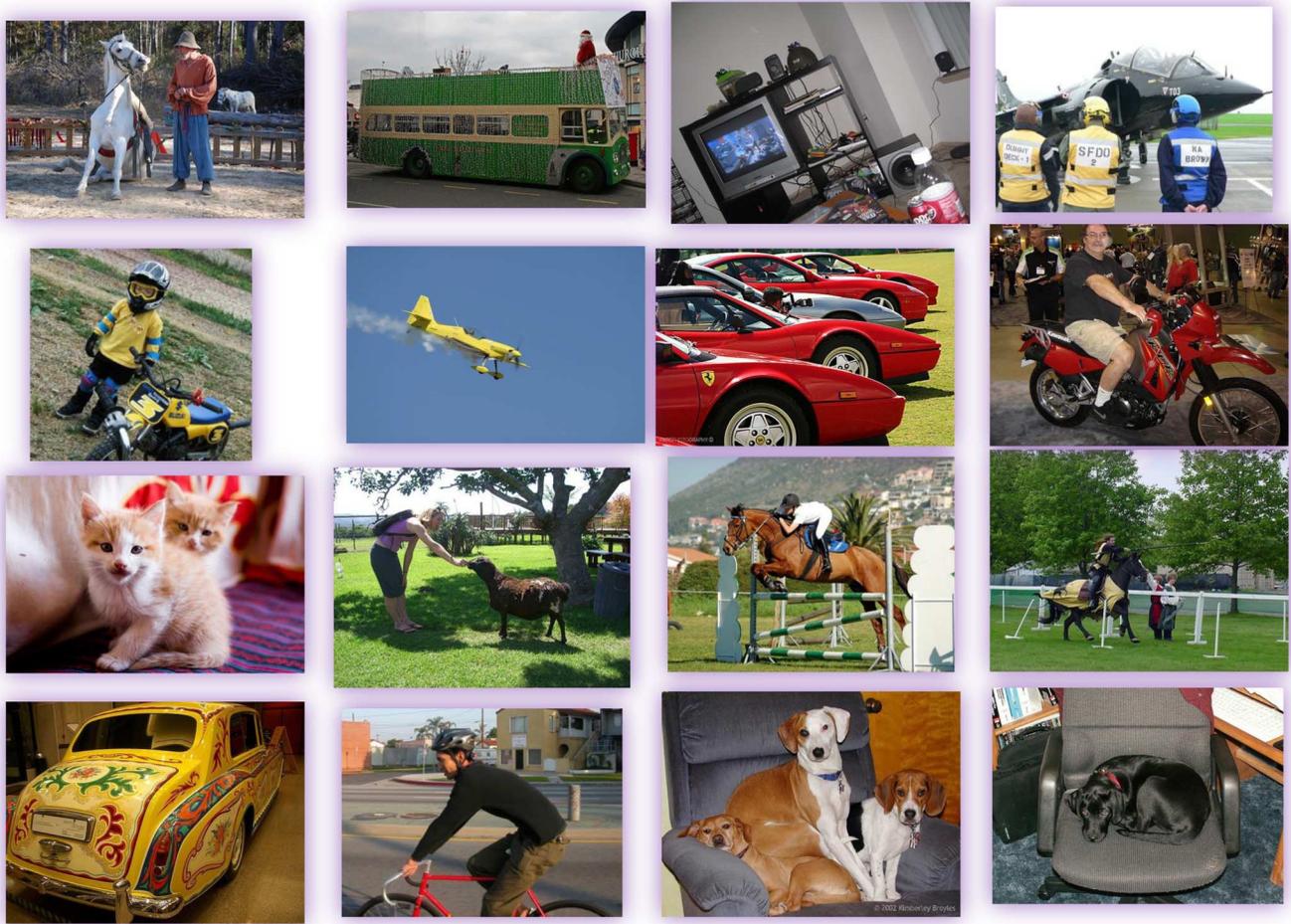


Fig. 3. Example images in the PASCAL VOC'07 dataset: the dataset contains 20 object classes (person, bird, cat, cow, dog, horse, sheep, aeroplane, bicycle, boat, bus, car, motorbike, train, bottle, chair, dining table, potted plant, sofa, and tv/monitor) from 4 groups (person, animal, vehicle and indoor).

the dataset were downloaded by querying for images of 20 common entities, including person, bird, cat, cow, dog, horse, sheep, aeroplane, bicycle, boat, bus, car, motorbike, train, bottle, chair, dining table, potted plant, sofa, and tv/monitor, from 4 big groups (person, animal, vehicle and indoor). Furthermore, the images were labeled carefully. There are 24,640 annotated objects in the dataset. This indicates that multiple objects from different classes may be presented in the same image. Each image has a corresponding annotation file giving information such as, the relevant filename, source, size, object name and boundary box of object and so on. We mainly use the object name information given by the annotation file during the training process in our experiments. Example images are given in Fig. 3. Since the procedure of feature extraction does not need the boundary box of object. Thus, the process of object detection is not included in our system for image annotation.

In particular, the training and test subsets contain 5,011 and 4,952 images, respectively. We further divide the training set into two subsets, one with 4500 images for training and one with 511 images for validation. We randomly conduct the partition 10 times to obtain error bar in performance evaluation. In the training set (4500 images), we use 10%, 20%, 30% 50%, 70% and 90% as labeled data and the rest as unlabeled data to conduct semi-supervised learning. To evaluate the supervised

learning algorithms, we do not use unlabeled data. For all methods, we tune the parameters on the set with 10% labeled data and 90% unlabeled data and use the tuned parameters in other settings.

C. GIST Descriptor

We use the GIST descriptor in our experiments. Originally, the word “gist” refers to an abstract representation of the scene that spontaneously activates memory representations of scene categories [41]. Therefore, the GIST descriptor is a kind of biologically-inspired features which uses image features from the visual cortex. In order to obtain the GIST descriptor, the images should be processed through three main low-level visual “channels”, including color channel, intensity channel and orientation channel, at multiple spatial scales. Within the channels, the operations called center-surround should be performed between different scales. The GIST descriptor can be received after linearly combining the center-surround results to yield a saliency map. In this Section, we use the GIST descriptor extracted by Guillaumin *et al.* [30].

D. Baselines and Performance Measures

In this Section, we conduct experiments to evaluate the effectiveness of Hessian regularized SVM. We compare it with

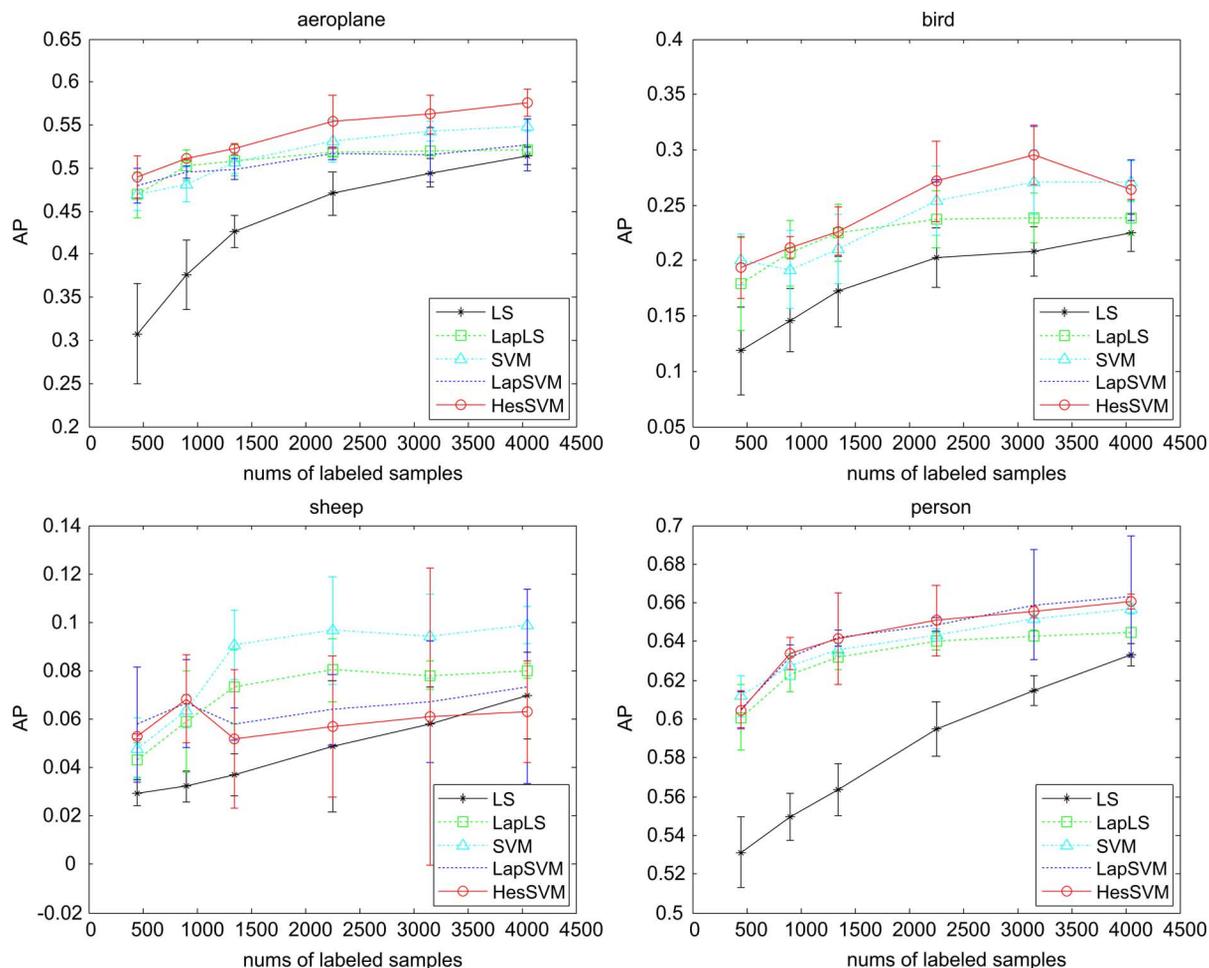


Fig. 4. We compared Hessian regularized SVM (HesSVM) with Laplacian regularized SVM (LapSVM), Laplacian regularized least squares (LapLS), SVM, and least squares (LS) on some representative objects, including aeroplane, bird, sheep and person. In each subfigure, the x-coordinate is the number of the labeled samples in the training set and y-coordinate is the average precision. These subfigures suggest the effectiveness of Hessian regularized SVM.

several popular methods including kernel least squares, SVM, Laplacian regularized kernel least squares, and Laplacian regularized SVM. For all methods, regularization parameters λ_K and λ_H (in Hessian regularized SVM, or the λ_I in Laplacian regularized SVM) are tuned from the candidate set $\{10^i | i = -6, -5, \dots, 5, 6\}$. The parameter k (the number of neighbors in k -nearest neighbors) used in computing the Hessian regularization is set to 100 for all experiments.

We use the average precision (AP) as the performance evaluation criterion for each class in Fig. 4 and the mean AP (mAP) over all classes in Fig. 5.

E. Experimental Results and Analysis

In Fig. 4, we compare the proposed Hessian regularized SVM with Laplacian regularized SVM, Laplacian regularized least squares (in RKHS), SVM, and least squares (in RKHS) on the PASCAL VOC'07 dataset with four representative categories. Each subfigure of this figure shows the performance curves of a particular category from aeroplane, bird, sheep and person. In each subfigure, the x-coordinate is the number of the labeled samples in the training set and y-coordinate is the average precision. It shows that the person class achieves the highest AP and the sheep class achieves the lowest AP. This is because the

training sample number of different category varies greatly, e.g., the samples size of the person category in the training subset is 2,050 and the samples size of the sheep category is only 96. The subfigures of bird and sheep show that the AP of Hessian regularized support vector machines appears to decrease when there are more labeled samples, because of the following three reasons: 1) the size of the training set for a particular category is small; 2) we conduct the partition of training set randomly; and 3) semi-supervised learning algorithms is sensitive to the distribution of unlabeled samples. By analyzing these subfigures, we can see that Hessian regularized SVM is superior to other baselines in general.

In Fig. 5, we show the mAP boxplots of different methods. There are six subfigures, each of which corresponds to the performance obtained by a particular number (450, 900, 1,350, 2,250, 3,150 and 4,050) of labeled training samples. For all subfigures, the number of training samples including labeled and unlabeled samples is 4,500. These boxplots show Hessian regularized SVM is superior to the baseline methods and validate the effectiveness of Hessian regularization for encoding the local geometry of the compact support of the marginal distribution. In addition, Fig. 5 shows the mAP appears to increase when there are more labeled samples.

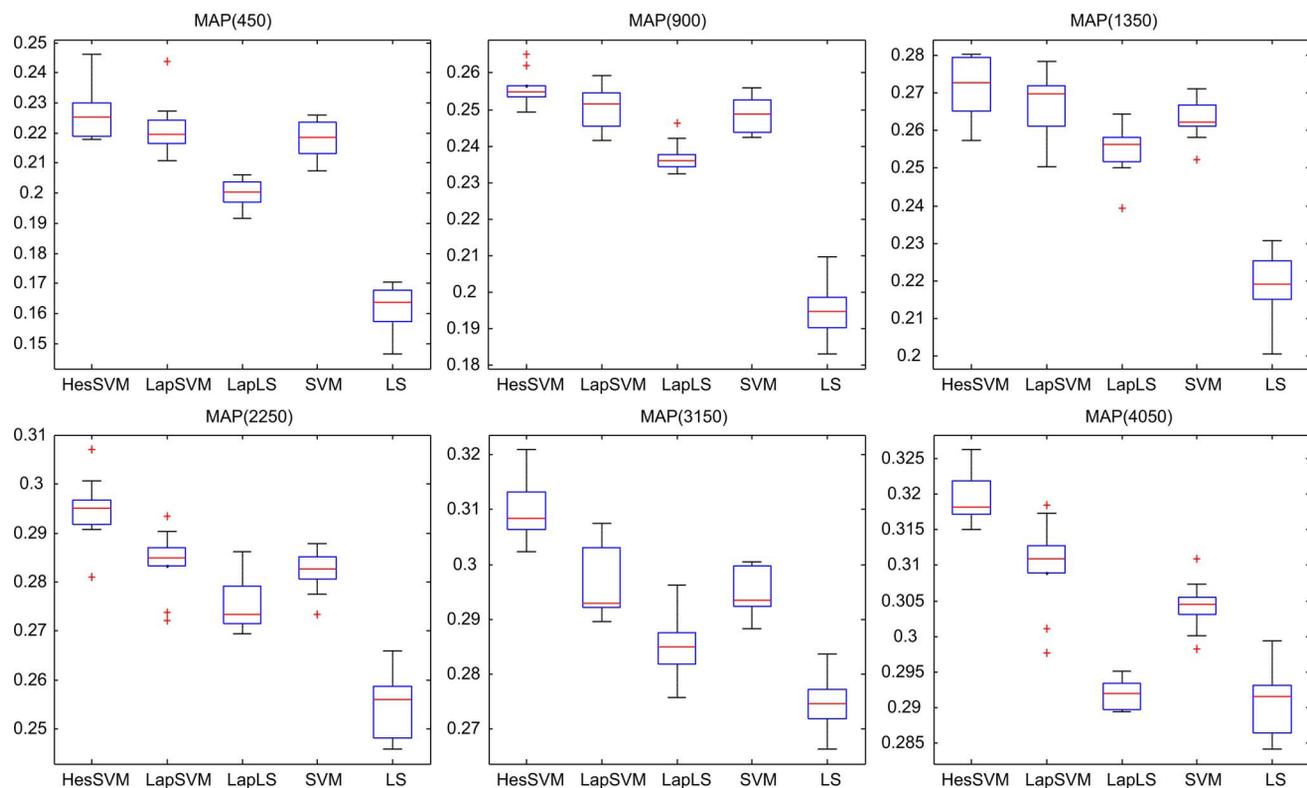


Fig. 5. The mAP boxplots of different methods. There are six subfigures, each of which corresponds to the performance obtained by a particular number (450, 900, 1,350, 2,250, 3,150 and 4,050) of labeled training samples. For all subfigures, the number of training samples including labeled and unlabeled samples is 4500.

V. CONCLUSIONS

This paper studies the image annotation problem on the cloud. The presented scheme includes image compression through Hamming compressed sensing and semi-supervised classification for semantic labeling. Technically, we develop Hessian regularized support vector machines that use a large number of unlabeled images to encode the local geometry of the compact support of the marginal distribution. The underlying assumption is that close image pairs induce similar conditional distribution pairs, and thus it is important to understand the local geometry of the unlabeled images. Although graph Laplacian was developed to explore the local geometry of a data manifold, it is not as good as the Hessian regularization induced from Hessian Eigenmaps. The nullspace of the graph Laplacian is the constant function along the compact support of the marginal distribution, while the nullspace of the Hessian regularization is spanned by the constant function and the original isometric coordinates. The nullspace is important because the classification function will not be penalized along the nullspace. The nullspace of the Hessian regularization is richer than that of the graph Laplacian, so it is superior to graph Laplacian for modeling the local geometry of the compact support of the marginal distribution. Thorough empirical studies confirmed the proposed learning model.

However, the success of Hessian regularized support vector machines for image annotation requires a precise estimation of the distribution of unlabeled samples. Thus a large number of unlabeled samples should be collected and the training sample size is larger than traditional supervised learning algorithms.

That means the proposed algorithm requires more computing and storage costs to achieve a high performance.

In the future, we will apply the proposed Hessian regularized support vector machines to other applications, e.g., text annotation, scene classification and spoken letter recognition. These applications will attract more attentions with the rapid development of mobile visual search and the cloud computing. In addition, Hessian regularized support vector machines have some important parameters that affect the algorithm performance, and thus the automatic selection of the parameters is important and should be investigated carefully.

REFERENCES

- [1] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A view of cloud computing," *Commun. ACM*, vol. 53, no. 4, pp. 50–48, 2010.
- [2] K. Barnard, P. Duygulu, N. de Freitas, D. A. Forsyth, D. M. Blei, and M. I. Jordan, "Matching words and pictures," *J. Mach. Learning Res.*, vol. 3, pp. 1107–1135, 2003.
- [3] M. Belkin and P. Niyogi, "Laplacian Eigenmaps and spectral techniques for embedding and clustering," *Adv. Neural Inf. Process. Syst.*, vol. 14, pp. 585–591, 2002.
- [4] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *J. Mach. Learning Res.*, vol. 7, pp. 2399–2434, 2006.
- [5] M. Bilenko, S. Basu, and R. J. Mooney, "Integrating constraints and metric learning in semi-supervised clustering," in *Proc. 21st Int. Conf. Mach. Learning*, 2004, p. 11.
- [6] D. M. Blei and J. D. McAuliffe, "Supervised topic models," *Adv. Neural Inf. Process. Syst.*, vol. 19, pp. 993–1022, 2007.
- [7] P. T. Boufounos and R. G. Baraniuk, "One-bit compressive sensing," in *Proc. Int. Conf. Inf. Sci. Syst.*, 2008, pp. 16–21.
- [8] J. Bresnahan, K. Keahey, D. Labissoniere, and T. Freeman, "Cumulus: An open source storage cloud for science," in *Proc. 2nd Int. Workshop Scientific Cloud Comput.*, 2011, pp. 25–32.

- [9] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.
- [10] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27–54, 2001.
- [11] Y. Chen, J. Z. Wang, and R. Krovetz, "Content-based image retrieval by clustering," in *Proc. ACM Int'l Conf. on Multimedia Inform. Retrieval*, 2003, pp. 193–200.
- [12] J. Cui, F. Wen, R. Xiao, Y. Tian, and X. Tang, "EasyAlbum: An interactive photo annotation system based on face clustering and re-ranking," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2007, pp. 367–376.
- [13] O. Chapelle, B. Scholkopf, and A. Zien, *Semi-Supervised Learning*. Cambridge, MA, USA: MIT, 2006, vol. 2.
- [14] V. Chandrasekhar, G. Takacs, D. M. Chen, S. S. Tsai, Y. A. Reznik, R. Grzeszczuk, and B. Girod, "Compressed histogram of gradients: A low-bitrate descriptor," *Int. J. Comput. Vision*, vol. 96, no. 3, pp. 384–399, 2012.
- [15] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [16] D. L. Donoho and C. Grimes, "Hessian Eigenmaps: New locally linear embedding techniques for high-dimensional data," *Proc. Nat. Acad. Sci. USA*, vol. 100, no. 10, pp. 5591–5596, 2003.
- [17] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and gabor features for gait recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1700–1715, Oct. 2007.
- [18] P. Duygulu, K. Barnard, N. de Freitas, and D. Forsyth, "Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary," in *Proc. Eur. Conf. Comput. Vision*, 2002, vol. 4, pp. 97–112.
- [19] D. Tao, X. Tang, X. Li, and Y. Rui, "Direct kernel biased discriminant analysis: A new content-based image retrieval relevance feedback algorithm," *IEEE Trans. Multimedia*, vol. 8, no. 4, pp. 716–727, Aug. 2006.
- [20] J. Eells and L. Lemaire, *Selected Topics in Harmonic Maps*. Providence, RI, USA: AMS, 1983.
- [21] M. Everingham, L. Van Gool, C. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes Challenge 2007," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, 2010.
- [22] I. Foster, "Service-oriented science," *Sci.*, vol. 308, no. 6, pp. 814–817, 2005.
- [23] Y. Gao, L. Jin, C. He, and G. Zhou, "Handwriting character recognition as a service: A new handwriting recognition system based on cloud computing," in *Proc. ICDAR*, 2011, pp. 885–889.
- [24] B. Geng, Y. Li, D. Tao, M. Wang, Z.-J. Zha, and C. Xu, "Parallel lasso for large-scale video concept detection," *IEEE Trans. Multimedia*, vol. 14, no. 1, pp. 55–65, Feb. 2012.
- [25] B. Geng, D. Tao, C. Xu, L. Yang, and X.-S. Hua, "Ensemble manifold regularization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 6, pp. 1227–1233, Jun. 2012.
- [26] B. Geng, L. Yang, C. Xu, and X.-S. Hua, "Content-aware ranking for visual search," in *Proc. IEEE Int. Conf. Comput. Vision Pattern Recogn.*, 2010, pp. 3400–3407.
- [27] B. Girod, V. Chandrasekhar, R. Grzeszczuk, and Y. A. Reznik, "Mobile visual search: Architectures, technologies, and the emerging mpeg standard," *IEEE MultiMedia*, vol. 18, no. 3, pp. 86–94, 2011.
- [28] P. H. Gosselin and M. Cord, "Active learning methods for interactive image retrieval," *IEEE Trans. Image Process.*, vol. 17, no. 7, pp. 1200–1211, Jul. 2008.
- [29] N. Guan, D. Tao, Z. Luo, and B. Yuan, "NeNMF: An optimal gradient method for nonnegative matrix factorization," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2882–2898, Jun. 2012.
- [30] M. Guillaumin, J. Verbeek, and C. Schmid, "Multimodal semisupervised learning for image classification," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recogn.*, 2010, pp. 902–909.
- [31] X. He, D. Cai, and J. Han, "Learning a maximum margin subspace for image retrieval," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 2, pp. 189–201, Apr. 2008.
- [32] T. Joachims, T. Finley, and C.-N. Yu, "Cutting-plane training of structural SVMs," *Mach. Learning*, vol. 77, no. 1, pp. 27–59, 2009.
- [33] D. Tao, X. Li, and S. J. Maybank, "Negative samples analysis in relevance feedback," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 4, pp. 568–580, Nov. 2007.
- [34] K.-I. Kim, F. Steinke, and M. Hein, "Semi-supervised regression using Hessian energy with an application to semi-supervised dimensionality reduction," *Adv. Neural Inf. Process. Syst.*, pp. 979–987, 2009.
- [35] Y.-Y. Lin, T.-L. Liu, and H.-T. Chen, "Semantic manifold learning for image retrieval," in *Proc. ACM Int. Conf. Multimedia*, 2005, pp. 249–258.
- [36] A. Losup, S. Ostermann, N. Yigitbasi, R. Prodan, T. Fahringer, and D. Epema, "Performance analysis of cloud computing services for many-tasks scientific computing," *IEEE Trans. Parallel Distrib. Syst.*, vol. 22, no. 6, pp. 931–945, 2011.
- [37] D. Tao, X. Tang, X. Li, and X. Wu, "Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 7, pp. 1088–1099, Jul. 2006.
- [38] P. Mell and T. Grance, "The NIST definition of Cloud Computing," Nat. Inst. of Standards Technol., 2011.
- [39] F. Monay and D. Gatica-Perez, "On image auto-annotation with latent space models," in *Proc. ACM Int. Conf. Multimedia*, 2003, pp. 275–278.
- [40] K. Morik, P. Brockhausen, and T. Joachims, "Combining statistical learning with a knowledge-based approach—A case study in intensive care monitoring," in *Proc. 16th Int. Conf. Mach. Learning*, 1999, pp. 268–277.
- [41] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vision*, vol. 42, no. 3, pp. 145–175, 2001.
- [42] D. Tao, X. Li, X. Wu, and S. J. Maybank, "Geometric mean for subspace selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 260–274, Feb. 2009.
- [43] S. Papadopoulos, C. Zigkolis, Y. Kompatsiaris, and A. Vakali, "Cluster-based landmark and event detection for tagged photo collections," *IEEE Multimedia*, vol. 18, no. 1, pp. 52–63, Feb. 2011.
- [44] J. R. Quinlan, "Induction of decision trees," *Mach. Learning*, vol. 1, no. 1, pp. 81–106, 1986.
- [45] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323–2326, 2000.
- [46] S. Shalev-Shwartz, Y. Singer, and N. Srebro, "Pegasos: Primal estimated sub-gradient solver for SVM," in *Proc. 24th Annu. Int. Conf. Mach. Learning*, 2007, pp. 807–814.
- [47] Y. Shao, Y. Zhou, X. H. D. Cai, and H. Bao, "Semi-supervised topic modeling for image annotation," in *Proc. ACM Int. Conf. Multimedia*, 2009, pp. 521–524.
- [48] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.
- [49] R. Shi, C.-H. Lee, and T.-S. Chua, "Enhancing image annotation by integrating concept ontology and text-based Bayesian learning model," in *Proc. ACM Int. Conf. Multimedia*, 2007, pp. 341–344.
- [50] M. Song, D. Tao, C. Chen, J. Bu, J. Luo, and C. Zhang, "Probabilistic exposure fusion," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 341–357, Jan. 2012.
- [51] M. Song, D. Tao, C. Chen, X. Li, and C. W. Chen, "Color to gray: Visual cue preservation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1537–1552, Aug. 2010.
- [52] F. Steinke and M. Hein, "Non-parametric regression between manifolds," *Adv. Neural Inf. Process. Syst.*, pp. 1–8, 2008.
- [53] J. Tenenbaum, V. Silva, and J. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, pp. 2319–2323, 2000.
- [54] X. Tian, D. Tao, and Y. Rui, "Sparse transfer learning for interactive video search reranking," *ACM Trans. Multimedia Comput., Commun., Applicat.*, vol. 8, no. 3, p. 26, 2012.
- [55] S. Tong and E. Y. Chang, "Support vector machine active learning for image retrieval," in *Proc. ACM Int. Conf. Multimedia*, 2001, pp. 107–118.
- [56] D. Tao and L. Jin, "Discriminative information preservation for face recognition," *Neurocomputing*, vol. 91, pp. 11–20, 2012.
- [57] V. Vapnik, *Statistical Learning Theory*. New York, NY, USA: Wiley-Interscience, 1998.
- [58] L. M. Vaquero and J. C. Morán, "The challenge of service level scalability for the cloud," *Int. J. CAC*, vol. 1, no. 1, pp. 34–44, 2011.
- [59] C. Wang, D. M. Blei, and Li F.-F., "Simultaneous image classification and annotation," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recogn.*, 2009, pp. 1903–1910.
- [60] Q. Zhang, L. Cheng, and R. Boutaba, "Clouding computing: State-of-the-art and research challenges," *J. Internet Service Applicat.*, vol. 1, no. 1, pp. 7–18, 2010.
- [61] T. Zhang, D. Tao, X. Li, and J. Yang, "Patch alignment for dimensionality reduction," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1299–1313, Dec. 2009.
- [62] T. Zhou and D. Tao, "Hamming Compressed Sensing," 2011, arXiv preprint, arXiv: 1110.0073.

- [63] T. Zhou, D. Tao, and X. Wu, "NESVM: A fast gradient method for support vector machines," in *Proc. IEEE Int. Conf. Data Mining*, 2010, pp. 679–688.
- [64] X. Zhu, *Semi-Supervised Learning Literature Survey*, *Computer Science*. Madison, WI, USA: Univ. Wisconsin-Madison, 2006.



Dapeng Tao received the B.S. degree in Electronics and Information Engineering from Northwestern Polytechnical University, Xi'an, China. He is currently a Ph.D. candidate in Information and Communication Engineering at South China University of Technology, Guangzhou, China. His research interests include machine learning and computer vision.



Lianwen Jin received a BS degree from the University of Science and Technology of China and a Ph.D. degree from South China University of Technology in 1991 and 1996, respectively. He is now a professor with the School of Electronic and Information Engineering, South China University of Technology. He is the author of more than 100 scientific papers. He is member of IEEE Signal Processing Society, IEEE Communication Society, and IEEE Computer Society, China Image and Graphics Society, the Cloud Computing Experts

Committee of China Institute of Communications. He received the award of New Century Excellent Talent Program of MOE in 2006, the Guangdong Pearl River Distinguished Professor award in 2011, respectively. He served as Program Committee member for a number of international conferences, including ICMLC 2007-2011, ICFHR2008-2012, ICDAR2009, ICDAR2013, ICPR2010, ICPR2012, ICMLA2012 *et al.* His research interests include image processing, handwriting analysis and recognition, machine learning, cloud computing, and intelligent systems.



Weifeng Liu is currently an Associate Professor with College of Information & Control Engineering, China University of Petroleum (East China). He received double B.S. degree (Automation and Business Administration) and Ph.D. degree (Pattern Recognition and Intelligent Systems) from University of Science and Technology of China (USTC) in 2002 and 2007 separately. He was a visiting scholar with Centre for Quantum Computation & Intelligent Systems (QCIS), Faculty of Engineering & IT, University of Technology Sydney (UTS) in Australia from September 2011 to August 2012. His currently research interests include Computer Vision, Pattern Recognition and Machine Learning.

Xuelong Li is a full professor with the Center for OPTical IMagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shaanxi, P. R. China.