10-2009

Distance Metric Learning from Uncertain Side Information with Application to Automated Photo Tagging

Lei WU
University of Science and Technology of China

Steven C. H. HOI
Singapore Management University, CHHOI@smu.edu.sg

Rong JIN
Michigan State University

Jianke ZHU
ETH Zurich

Nenghai YU
MOE-Microsoft Keynote Lab of MCC, Hefei, China

DOI: https://doi.org/10.1145/1631272.1631293

Citation

WU, Lei; HOI, Steven C. H.; JIN, Rong; ZHU, Jianke; and YU, Nenghai. Distance Metric Learning from Uncertain Side Information with Application to Automated Photo Tagging. (2009). *MM'09: Proceedings of the 2009 ACM International Conference on Multimedia and co-located Workshops, October 19-24, 2009, Beijing, China*. 135-144. Research Collection School Of Information Systems.
Available at: https://ink.library.smu.edu.sg/sis_research/2369

Distance Metric Learning from Uncertain Side Information with Application to Automated Photo Tagging

Lei Wu[♮†][*], Steven C.H. Hoi[†], Rong Jin[♯], Jianke Zhu[‡], and Nenghai Yu[♮]
[†]School of Computer Engineering, Nanyang Technological University;
[♮]MOE-MS Keynote Lab of MCC,University of Science and Technology of China;
[♯]Dept. of Computer Sci. and Eng., Michigan State University; [‡]ETH Zurich
leiwu@live.com, chhoi@ntu.edu.sg, rongjin@cse.msu.edu, zhu@vision.ee.ethz.ch,
ynh@ustc.edu.cn

ABSTRACT

Automated photo tagging is essential to make massive unlabeled photos searchable by text search engines. Conventional image annotation approaches, though working reasonably well on small testbeds, are either computationally expensive or inaccurate when dealing with large-scale photo tagging. Recently, with the popularity of social networking websites, we observe a massive number of user-tagged images, referred to as "*social images*", that are available on the web. Unlike traditional web images, social images often contain tags and other user-generated content, which offer a new opportunity to resolve some long-standing challenges in multimedia. In this work, we aim to address the challenge of large-scale automated photo tagging by exploring the social images. We present a retrieval based approach for automated photo tagging. To tag a test image, the proposed approach first retrieves k social images that share the largest visual similarity with the test image. The tags of the test image are then derived based on the tagging of the similar images. Due to the well-known semantic gap issue, a regular Euclidean distance-based retrieval method often fails to find semantically relevant images. To address the challenge of semantic gap, we propose a novel *probabilistic distance metric learning* scheme that (1) automatically derives constraints from the uncertain side information, and (2) efficiently learns a distance metric from the derived constraints. We apply the proposed technique to automated photo tagging tasks based on a social image testbed with over 200,000 images crawled from Flickr. Encouraging results show that the proposed technique is effective and promising for automated photo tagging.

MM'09, October 19–24, 2009, Beijing, China.
Copyright 2009 ACM 978-1-60558-608-3/09/10 ...$10.00.

Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing; I.2.6 [**Artificial Intelligence**]: Learning; I.4.7 [**Image Processing and Computer Vision**]: Feature Measurement

General Terms

Algorithm, Experimentation

Keywords

automated photo tagging, distance metric learning, uncertain side information

1. INTRODUCTION

Due to the popularity of digital cameras, digital photos can be easily created in our daily life. The massive unlabeled photos have posed a huge challenge for image retrieval tasks. One solution is to automatically annotate images with keywords or social tags. With the auto-annotations, an image retrieval problem is converted into a text retrieval problem, which enjoys both efficient computation and high retrieval accuracy.

In general, the objective of an automated image annotation task is to assign a set of semantic labels or tags to a novel image, based on some pre-trained models. A conventional approach usually consists of two steps: (1) extracting visual features for image representation [19], and (2) building classification models from a collection of manually-labeled training data [3]. In literature, numerous studies have been devoted to automated image annotation and object recognition tasks [17, 24].

Despite encouraging results in recent years, conventional image annotation approaches, which usually work well on small-sized testbeds with high quality labels, often fail to handle large scale real world photo tagging applications. One major challenge faced by large-scale photo annotation is primarily due to the well-known semantic gap between low-level features and high-level semantic concepts. Besides, it is also expensive and time-consuming to collect a large set of manually-labeled training data in the conventional methods. Hence, it has become an urgent need to develop new paradigms for automated photo tagging beyond the conventional approaches.

[*]This work was performed when Mr. Lei Wu was a research assistant at Nanyang Technological University.

Recently, with the popularity of social networking websites, we have witnessed the generation of massive user-tagged images on the web, which we refer to as "*social images*". Unlike traditional web images, social images often contain tags and rich user-generated content, which offer a new opportunity to resolve some long-standing challenges in multimedia, for instance the semantic gap. In this paper, we investigate an emerging retrieval-based paradigm [29] for automated photo tagging by mining massive social images freely available on the web. The basic idea of the retrieval-based paradigm is to first retrieve a set of k most similar images for a test photo from the social image repository, and then to assign the test photo with a set of t most relevant tags associated with the set of k retrieved social images.

The key of the retrieval-based photo tagging paradigm is to *accurately* identify and retrieve the set of top k (semantically) similar photos, which generally relies on two key components: (1) a feature representation scheme to extract salient visual features, and (2) a distance measure method to effectively calculate distances for the extracted features. In this paper, we focus our main efforts on tackling the second challenge. In particular, by assuming features are represented in vector space, our goal is to learn an optimal distance metric for distance measure, which is often known as "distance metric learning" (DML) [32].

Many studies have been devoted to DML due to its importance for many applications. Existing DML studies often assume the learning task is provided with explicit *side information* given in the form of either class labels [30, 16] or pairwise constraints [32, 1] where each pairwise constraint indicates whether two examples are similar ("must-link") or dissimilar ("cannot-link"). The side information can be collected from users in some environments, such as relevance feedback log in CBIR [13]. Besides the explicit side information, regular DML studies usually assume *perfect* side information. Such assumptions make regular DML techniques difficult to be applied in our web application. This is because in our case, most images are labeled by a number of tags (some of them may be noisy). As a result, we often find a partial overlap between two images in their assigned tags, which makes it difficult to decide if two images form a must-link constraint. The side information derived from the tags and other rich content of social images, referred to as *uncertain side information*, leads to a new challenge in DML as opposed to the conventional set where "hard" side information is available.

To this end, this paper presents a novel *probabilistic distance metric learning* (PDML) framework, which aims to learn effective metrics from uncertain side information with application to automated photo tagging. In general, the proposed framework consists of two steps: (1) a graphical model learning approach to discover probabilistic side information from hidden side information contained implicitly in rich user-generated content of social image data; and (2) a probabilistic metric learning method to find an optimal distance metric from probabilistic side information. To the best of our knowledge, this is the first probabilistic approach to learn an optimal metric from uncertain side information.

As a summary, the key contributions of this paper include: (1) a novel probabilistic DML framework to learn distance metrics from uncertain side information; (2) an effective algorithm, i.e., probabilistic Relevant Component Analysis (pRCA), to learn an optimal metric from probabilistic side information; (3) a new solution using the PDML technique to an emerging important application, i.e., automated photo tagging; (4) extensive experiments to compare our method with a number of state-of-the-art DML algorithms, in which very encouraging results were obtained.

The rest of this paper is organized as follows. Section 2 reviews related work and background. Section 3 presents an overview of our probabilistic DML framework. Section 4 presents a graphical model approach to find probabilistic side information from a social image repository. Section 5 proposes an efficient algorithm to learn distance metrics from probabilistic side information. Section 6 discusses an application of our technique for exploring the social image repository in automated photo tagging tasks. Section 7 discusses experimental results, Section 8 discusses limitations of our work, and finally Section 9 concludes this work.

2. RELATED WORK AND BACKGROUND

Our work is mainly related to two groups of research. One group is the work on exploring web images and photos for automated image/photo annotation and object recognition [21, 26, 33, 31]. The other group is the work related to distance metric learning (DML) research [32, 1, 23, 6]. Due to limited space, we briefly review some most representative and relevant studies in both sides.

2.1 Automated Photo Tagging

Our work is related to automated image/photo annotation that has been actively studied over the past decade in multimedia community. Among a variety of conventional approaches, a widely-studied paradigm is the supervised classification approach, in which classification models, such as SVM [8], are trained from a collection of human-labeled training data for a set of predefined semantic concept/object categories [3, 4, 7, 28]. Besides, semi-supervised learning methods are also explored in recent literature [18, 12].

Recently, there is a surge of emerging interests in exploring web photo repositories for image annotation/object recognition problems. One promising approach is the retrieval-based (or termed "search-based") paradigm [21, 29, 26, 27]. Russell et al. [21] built a large collection of web images with ground truth labels for helping object recognition research. Wang et al. [29] proposed a fast search-based approach for image annotation by some efficient hashing technique. Rege et al. [20] utilized visual and text modalities simultaneously in clustering images. Wen-Yen et al.[5] proposed the combinational collaborative filtering model for personalized community recommendation. Torralba et al. [26] proposed efficient image search and scene matching techniques for exploring a large-scale web image repository. These work usually concerned more on fast indexing and search techniques, while we focus on learning more effective distance metrics. Yan et al. [33] investigated a learning based method for improving the efficiency of manual image annotation with the hybrid of tagging and borrowing. Different from their work, we investigate fully automatic photo annotation, which also can be extended to help manual image annotation. In addition, we could also apply our effective distance metric learning and automatic photo annotation techniques to facilitate some emerging applications in computer graphics, such as image completion and inpainting by exploring web photo repositories [11, 25].

2.2 Distance Metric Learning

From a machine learning point of view, our work is closely related to DML studies. Firstly, we review some basics of DML. Given a set of n data examples $X = \{x_i \in \mathbb{R}^d\}_{i=1}^n$ in d-dimensional vector space, the Mahalanobis distance between any two examples x_i and x_j is defined as:

$$d_M(x_i, x_j) = \sqrt{((x_i - x_j)^\top M(x_i - x_j))} \tag{1}$$

where M is a positive semi-definite matrix that satisfies the property of valid metric and can be decomposed as $M = A^\top A$. The goal of DML is to find an optimal Mahalanobis metric M from training data (side information) that can be either class labels or general pairwise constraints [32].

DML can be roughly divided into two major categories. One is to learn metrics with explicit class labels, such as Neighbourhood Components Analysis (NCA) [16], which are often used for classification [9, 10, 30, 34]. The other is to learn metrics from pairwise constraints for clustering and retrieval. Examples include Relevance Component Analysis (RCA) [1] and Discriminative Component Analysis (DCA) [15], amongst others [32, 14]. Our work is more related to the second category, though some methods in the former category could be converted to the latter.

Unlike most existing DML methods that assume explicit side information is provided in the form of either class labels or pairwise constraints, in our DML problem, no explicit side information is directly given for the learning task. Instead, our goal is to learn metrics from uncertain side information, which is hidden in the rich contents of social image training data in our application.

2.3 Relevant Component Analysis

Here we review a well-known and effective DML technique, i.e., Relevant Component Analysis (RCA) [1]. The basic idea of RCA is to identify and down-scale global unwanted variability within the data. In particular, RCA suggests to change the feature space used for data representation by a global linear transformation in which relevant dimensions are assigned with large weights. More formally, given a set of data examples $X = \{x_i\}_{i=1}^n$ and a collection of pairwise constraints indicating whether two data examples are similar (or dissimilar). RCA forms a set of m "**chunklets**" $C_j = \{x_{ji}\}_{i=1}^{n_j}$ where $j = 1, \ldots, m$. Each *chunklet* is defined as a group of data examples linked together by similar pairwise constraints ("must-link").

The optimal transformation by RCA is then computed as $A = \hat{C}^{-1/2}$ and the Mahalanobis matrix is equal to the inverse of the average covariance matrix of chunklets, i.e., $M = \hat{C}^{-1}$, where $\hat{C}$ is defined as follows:

$$\hat{C} = \frac{1}{n} \sum_{j=1}^{m} \sum_{i=1}^{m_j} (x_{ji} - \mu_j)(x_{ji} - \mu_j)^T \tag{2}$$

where μ_j denotes the mean of j-th chunklet, x_{ji} denotes the i-th example in the j-th chunklet and n is the total number of examples. RCA enjoys a number of merits, such as being sound in theory, simple, efficient, and easy to implement. Similar to other conventional DML techniques, RCA also requires a set of similar pairwise constraints explicitly provided for the learning task, which thus cannot be directly applied in our problem unless side information can be discovered/provided. In this paper, we extend RCA techniques to resolve the DML task from uncertain side information.

3. METRIC LEARNING FRAMEWORK FOR AUTOMATED PHOTO TAGGING

We first give an overview of the proposed semantic metric learning framework for learning metrics from social image data. Figure 1 shows a flowchart illustrating the proposed framework with application to automated photo tagging.

In the figure, the right panel shows a retrieval-based photo tagging solution. Specifically, given a novel photo, the idea of the retrieval-based tagging approach is to firstly perform a similarity search for finding top k most similar photos from the social photo repository, and then annotate the novel photo with top t ranked tags associated with the k retrieved photos. Our main effort focuses on learning an effective metric to reduce semantic gap for the similarity search process, which is shown in the left panel of the flowchart. Below we discuss the main ideas of our metric learning framework.

Since no explicit side information is available, we cannot directly apply regular DML techniques. Hence, the first step towards DML is to discover *possible* side information from training data, which is essential to DML. In other words, we wish to find some forms of side information, which could indicate how likely two social images are similar or dissimilar. One solution is to discover some "**chunklets**" (similar to RCA) from training data such that images in the same chunklets are similar to each other, and images in different chunklets could be similar or dissimilar, up to the similarity of the two associated chunklets. Since such chunklets are not explicitly available (also cannot be easily formed as RCA), we refer them to as "*latent chuklets*". Intuitively, a latent chunklet can be viewed as a common semantic topic shared by the social images in the chunklet. Thus, it is possible that one image belongs to multiple chunklets.

To find the latent chunklets effectively and precisely, we propose a graphical model approach to estimate the probabilities of an image belonging to the chunklets. We refer to this step as "Latent Chunklet Estimation" (LCE). By LCE, we can obtain side information in the form of latent chunklets with probabilistic assignments, which we refer to as "probabilistic side information" or "uncertain side information". Finally, the last step of our semantic metric learning is to find an optimal metric from the probabilistic side information output by the graphical model approach. In this paper, we propose a new probabilistic relevant component analysis (pRCA) to solve this key task effectively.

Next we first present the LCE process followed by the proposed pRCA method in the subsequent section.

4. LATENT CHUNKLET ESTIMATION FOR SOCIAL IMAGE MODELING

Typically a social image contains rich information, such as tags, title, description, comments, visual content, etc. In this paper, we propose a graphical model approach to discover side information of latent chunklets from rich contents of social images. For simplicity, we focus on exploring two key types of information, i.e., textual and visual. It is not difficult to engage additional information in our framework.

4.1 Latent Chunklet Definition

First of all, we assume that there are m latent chunklets available, each of them represents a hidden topic z_i, in which both visual images and associated textual metadata (e.g. tags) in the chunklets are generated from the hidden topic.

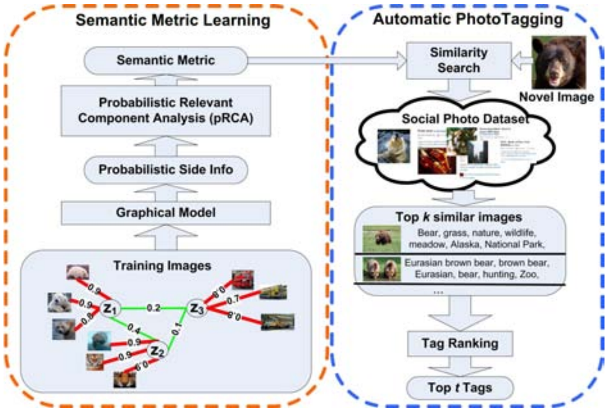


Figure 1: Flowchart illustrating the proposed metric learning framework for automated photo tagging

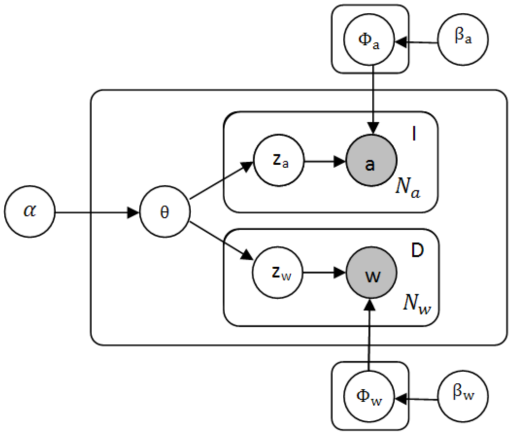


Figure 2: Graphical model for social image modeling

Figure 2 shows the graphical model for social image modeling. The upper part of the graph represents the visual model. The images can be represented by some local feature descriptor, e.g. bag of visual words representation [19], and each visual word a is generated from certain topic z_a by a multinomial distribution ϕ_a^z. In the left side, θ is a Dirichlet distribution with hyper parameter α. The lower part of the graph represents the textual model generating textual tags, in which w represents the tags. β is the parameter of the uniform Dirichlet prior on the per-topic word distribution, and α is the parameter of the uniform Dirichlet prior on the per-document topic distributions. For simplicity, we also assume that the tags are generated from a multinomial distribution ϕ_w^z parameterized by the topic z_w. Thus, a topic z contains two parts, i.e., $z = [z_a, z_w]$.

Our goal is to estimate the hidden distribution $P(z_a|I)$, the probability of an image I belonging to a certain topic z_a, and the hidden distribution $P(z_w|d)$, the probability of topic z_w existing in tag document d. Such conditional probabilities will be further used to predict the inter chunklet variation and intra chunklet variation. We discuss the generating process of the graphical model below.

Firstly, θ is the parameter for the topic distribution, which follows a Dirichlet distribution with parameter α:

$$\theta|\alpha \sim Dir(\alpha) \tag{3}$$

Further, given θ, topic z is drawn from a multinomial distribution, and Φ_a and Φ_w follow some Dirichlet distributions:

$$z|\theta \sim Multi(\theta), \quad \Phi_a|\beta_a \sim Dir(\beta_a), \quad \Phi_w|\beta_w \sim Dir(\beta_w) \tag{4}$$

Here we denote $\beta = [\beta_a, \beta_w]$. Finally, given topic z, both tags and visual words follow multinomial distributions:

$$w|z_w, \Phi_w \sim Multi(\phi_w^z), \quad a|z_a, \Phi_a \sim Multi(\phi_a^z) \tag{5}$$

4.2 Inferences

The main idea of the graphical model is to capture the conditional joint probability of tag document d and image x. A tag document is modeled by a bag of words $d = \{w\}$, and the image x is represented by a bag of visual words $x = \{a\}$. The joint probability $P(z, x, d|\alpha, \beta)$ can be written as:

$$P(z, x, d|\alpha, \beta) = \prod_{a,w} P(z, a, w|\alpha, \beta) = \prod_{a,w} \int_\theta P(z, a, w, \theta|\alpha, \beta)d\theta$$

where a represents a visual word in the social image, and w represents one of the tags with the social image. Further, according to the assumptions, the conditional joint probability of topic z, visual word a, tag w with respect to parameters α, β can be expressed as follows:

$$P(z, a, w, \theta|\alpha, \beta_a, \beta_w) \propto$$
$$P(w|z_w, \Phi_w)P(a|z_a, \Phi_a)P(z|\theta)P(\Phi_a|\beta_a)P(\Phi_w|\beta_w)$$

To calculate the chain of conditional probability in the above equation, Gibbs sampling is adopted. Although variational methods can also be used, we choose the Gibbs sampling for its simplicity and applicability to our problem. Specifically, it repeatedly draws a topic z with respect to the conditional distribution. Then visual words and tags are generated with the conditional probability given the topic z.

The objective of inference in the Gibbs sampling is to obtain the conditional distribution of hidden topic given the observed data. The Bayesian estimation of conditional distributions of tag, visual words, and topics are calculated as:

$$P(z_{w,i} = j|w) \propto \frac{n^w_{-i,j} + \beta_w}{n^{\cdot}_{-i,j} + W\beta_w}, \quad P(z_{a,i} = j|a) \propto \frac{n^a_{-i,j} + \beta_a}{n^{\cdot}_{-i,j} + A\beta_a}$$

$$P(x|z_{a,i} = j) \propto \frac{n^x_{-i,j} + \alpha}{n^x_{-i,\cdot} + m\alpha}, \quad P(d|z_{w,i} = j) \propto \frac{n^d_{-i,j} + \alpha}{n^d_{-i,\cdot} + m\alpha}$$

where $z_{w,i}$ represents topic z for tag w in the i^{th} sampling, $z_{a,i}$ denotes topic z for visual word a in the i^{th} sampling, and $n^w_{-i,j}$ is the frequency of tag w assigned to the j^{th} topic before the i^{th} sampling (and others have similar meanings). Besides, W is the size of the tag dictionary, A is the size of the visual word dictionary, and m is the number of topics.

With the above estimations, we can calculate the marginal by integrating out the parameter θ and sampling the topic with the distribution below:

$$P(z_{w,i} = j|z_{w,-i}, w) \propto \frac{n^w_{-i,j} + \beta_w}{n^{\cdot}_{-i,j} + W\beta_w} \times \frac{n^d_{-i,j} + \alpha}{n^d_{-i,\cdot} + m\alpha}$$

$$P(z_{a,i} = j|z_{a,-i}, a) \propto \frac{n^a_{-i,j} + \beta_a}{n^{\cdot}_{-i,j} + A\beta_a} \times \frac{n^x_{-i,j} + \alpha}{n^x_{-i,\cdot} + m\alpha}$$

Finally, we can calculate the topic relationship given parameter α and β as follows:

$$P(z_i, z_j|\alpha, \beta) \propto \frac{1}{N^2} \sum_{k=1}^{N} P(z_i, x_k, d_k|\alpha, \beta) P(z_j, x_k, d_k|\alpha, \beta)$$

Here we assume both z_w or z_a are sampled from a large latent chunklet set Z. z_i and z_j are any two topics from the set. As a summary, each topic z_i represents a chunklet. we can compute the conditional probability $P(z_i|x, d)$ that represents the relationship between the example and the chunklet, and the joint probability $P(z_i, z_j|\alpha, \beta)$ that represents the relationship between the two chunklets. These probabilities can be adopted and explored for DML.

5. PROBABILISTIC DML METHOD

5.1 Problem Definition

In this section, we present a probabilistic DML (PDML) method for learning metrics from probabilistic side information. Unlike regular RCA learning, the latent chunklets are represented by some probabilistic distributions rather than "strictly-hard" pairwise constraints. Therefore, the challenge of PDML is how to exploit the uncertain side information for optimizing the metric in the most effective way. Below we present a probabilistic RCA technique, which extends the regular RCA in a probabilistic metric learning approach. We first introduce some definitions and notations below.

Let us denote by x_i a d-dimensional visual feature vector of an image, and z_k one of m latent chunklets. Further, we denote by μ_k a center (mean) for a latent chunklet z_k, and $\mu = (\mu_1, \ldots, \mu_m)$ a matrix of all centers. Moreover, we denote by matrix $P = (p_1, \ldots, p_n)$ the membership probabilities of associating examples with chunklets, where $p_i = (p_i^{(1)}, \ldots, p_i^{(m)})$ is the probability distribution for the i-th example and $p_i^{(k)}$ represents the probability of observing example x_i given chunklet z_k, i.e., $p_i^{(k)} = p(x_i|z_k)$. In our approach, we initialize P by a prior probability matrix $P_0 = [p(x_i|z_k)]_{n\times m}$, which were obtained from LCE.

5.2 Probabilistic RCA

The objective of our DML task is to learn an optimal metric M in a d-dimensional feature vector space, i.e., $M \in \mathbb{R}^{d\times d}$. To exploit latent chunklets in DML, we formulate a probabilistic extension of RCA, termed as "Probabilistic Relevance Component Analysis" (pRCA), as follows:

$$\min_{M\succeq 0,\mu,P} \quad \sum_{i=1}^{n}\sum_{k=1}^{m} p_i^{(k)} \|x_i - \mu_k\|_M^2 - \lambda \log|M| \qquad (6)$$

$$s.t. \quad \|P - P_0\|_F^2 \le \gamma, \qquad (7)$$

$$\sum_k p_i^{(k)} = 1, p_i^{(k)} \ge 0, i = 1, \ldots, n \qquad (8)$$

where parameter $\gamma \ge 0$ constraints the difference between the prior probability matrix P_0 (known from LCE) and the proxy probability matrix P (unknown), λ is a regularization constant, and $\|\cdot\|_F$ denotes the Frobenius norm of a matrix.

The above formulation can be interpreted as a robust optimization problem with bounded uncertainty on the probability matrix P. In particular, for the objective function, the first term is to minimize the sum of squared distances from examples to their chunklet centers, and the second term is to prevent the solution M from being obtained by shrinking the entire solution space. For the constraints, the one in (7) is to restrict the matrix of desired probability assignments P without deviating too far from the prior matrix P_0, and the remaining set of constraints in (8) are used to enforce the probability requirements. The following corollary shows that RCA can be viewed as a special case of pRCA.

COROLLARY 1. *For the optimization in (6), when fixing the means of chunklets μ and the matrix of probability assignments P (assuming with hard assignments of 0 and 1), the pRCA formulation reduces to regular RCA learning.*

The proof of Corollary 1 can be found in Appendix A.

5.3 Algorithm

We now discuss techniques to solve the optimization of pRCA. Generally, the problem in (6) is a nonlinear optimization task containing three sets of variables M, P, and μ, where μ can be easily computed once P is found. It is often hard to solve the problem with global optima directly. To address this challenge, we present an iterative optimization algorithm by applying alternating optimization techniques [2], which is widely used to solve multi-variable nonlinear optimization tasks.

Our iterative optimization algorithm consists of three steps: (1) fixing P and μ to optimize M; (2) fixing M and μ to optimize P; and (3) fixing P and M to find μ. According to Corollary 1, the first step is equivalent to solving regular RCA, i.e., $M = \frac{1}{\lambda}\tilde{C}^{-1}$, where $\tilde{C}$ is the average chunklet covariance matrix with the given P. The last step is straightforward, i.e., $\mu = P^\top X$, where X is a matrix of all training data. We now focus on the second step. In particular, by fixing M and μ, the optimization can be rewritten as follows:

$$\min_{P} \quad \sum_{i=1}^{n}\sum_{k=1}^{m} p_i^{(k)} \|x_i - \mu_k\|_M^2 + \frac{\gamma}{2}\|P - P_0\|_F^2 \qquad (9)$$

$$s.t. \quad \sum_k p_i^{(k)} = 1, p_i^{(k)} \ge 0, i = 1, \ldots, n$$

where the constraint in (7) was moved to the objective. The above problem is a quadratic program (QP), which can be solved by some existing convex optimization software. However, for a real web application, the training data size can be very large, this poses a challenge of huge computation when solving a large-scale QP problem by a standard QP solver. To this end, we develop a fast algorithm, which is able to solve the above optimization efficiently.

To ease discussions, we notice all p_i's are completely decoupled in (9) given μ_k. Thus, we can rewrite (9) into a set of n independent optimization tasks, one for each p_i, i.e.,

$$\min_{p\in\mathbb{R}^m} \quad \sum_{k=1}^{m} p_k \|x_i - \mu_k\|_M^2 + \frac{\gamma}{2}\|p - p_0\|_2^2 \qquad (10)$$

$$s.t. \quad \sum_{k=1}^{m} p_k = 1, p_k \ge 0, k = 1, \ldots, m$$

It can be easily shown that solving the above problem is equivalent to solving the problem in (9). We now discuss a fast algorithm to solve this problem. We first introduce the Lagrangian of the optimization as follows:

$$\mathcal{L} = f^\top p + \frac{\gamma}{2}\|p - p_0\|_2^2 + \rho\Big(\sum_k p_k - 1\Big) - \eta \cdot p \qquad (11)$$

Algorithm 1 Probabilistic RCA Algorithm (pRCA)

```
INPUT:
    • training data matrix: $X \in \mathbb{R}^{n\times d}$
    • chunklet assignment probabilities: $P_0 \in \mathbb{R}^{n\times m}$
    • penalty parameter: $\gamma \geq 0$
OUTPUT:
    • optimized distance metric: $M^*$
initialize $P = P_0$, and $\mu = P^\top X$
repeat
  (1) compute $M$ by the following formula:
  $M = \big(\sum_{k=1}^m \sum_{i=1}^n p_i^k (x_i - \mu_k)(x_i - \mu_k)^\top\big)^{-1}$
  (2) find $P$ by solving QP problem in (9) as follows:
  for $i = 1$ to $n$ do
    $\mathbf{f}^\top = (\|x_i - \mu_1\|_M^2, \ldots, \|x_i - \mu_m\|_M^2)$
    $\mathbf{f} = \mathbf{sort}(\mathbf{f}, \text{'}descending\text{'})$
    find $\rho$ by Proposition 1
    for $k = 1$ to $m$ do
      $p_i^{(k)} = \max\Big(0, p_{0k} - \frac{1}{\gamma}(\rho + f_k)\Big)$
    end for
  end for
  (3) update the chunklet means: $\mu = P^\top X$
until convergence
```

where $f^\top = (\|x_i - \mu_1\|_M^2, \ldots, \|x_i - \mu_m\|_M^2)$, ρ is a Lagrange multiplier and η is a vector of non-negative Lagrange multipliers. By differentiating it with respect to p_k, we can get the following optimality condition:

$$\frac{\partial \mathcal{L}}{\partial p_k} = f_k + \gamma(p_k - p_{0k}) + \rho - \eta_k = 0$$

By applying the KKT condition, whenever $p_k > 0$, η_k should be zero. Therefore, if $p_k > 0$, we have the following result:

$$p_k = p_{0k} - \frac{1}{\gamma}(\rho + f_k)$$

Combining the fact that $p_k \geq 0$, we have the following:

$$p_k = \max\big(0, p_{0k} - \frac{1}{\gamma}(\rho + f_k)\big) \tag{12}$$

The next issue is to find the optimal ρ. The following proposition provides a solution to find the optimal value of ρ by a simple sorting approach.

PROPOSITION 1. *Let f' denote the vector by sorting f in decreasing order, the optimal value of ρ to the solution in (12) can be computed as: $\rho = -\frac{1}{\tau}\Big(\sum_{k=1}^{\tau}(f'_k - \gamma p_{0k}) + \gamma\Big)$, where τ can be found through a sorting approach, i.e.,*

$$\tau = \max_{k\in[1,n]} \Big\{k : f'_k - \frac{1}{k}\big(\sum_{j=1}^{k}(f'_j - \gamma p_{0k}) + \gamma\big) > 0\Big\} \tag{13}$$

By Proposition 1, we can solve the QP problem (10) in $\mathcal{O}(n\log(n))$, which is significantly faster than standard QP solvers with interior point methods that usually require $\mathcal{O}(n^3)$ complexity. Finally, we summarize the pseudo-code of the pRCA algorithm in Algorithm 1. The following corollary guarantees the convergence of the proposed algorithm.

COROLLARY 2. *Algorithm 1 converges to the local optimum for the optimization problem of probabilistic relevance component analysis in (6).*

It is not difficult to verify the above corollary by following the convergence theory of alternating optimization [2].

One of the advantages of pRCA is its robust to missing tagged images. In original RCA, the constraints are generated manually, there should be tags indicate which images should be in the same chunklet. In pRCA the probabilistic constraints as well as the chunklets are generated automatically by a graphical model based on their appearance features. Thus the model is more robust and automatic.

6. AUTOMATED PHOTO TAGGING

In this section, we discuss the application of pRCA to the exploitation of social photo repositories for automated photo tagging tasks. Given a novel photo, the automated tagging task is to annotate the photo labels or tags, which often reflect certain semantic concepts/objects. To overcome the limitation of conventional approaches, we investigate a retrieval based approach to automated photo tagging tasks by exploring a huge number of social photos freely available on the web. We formally formulate our approach as follows.

Let $I_q = \{x_q, \mathcal{T}_q\}$ denote a query image for tagging, where x_q represents the visual contents of the image, and $\mathcal{T}_q$ denotes a set of unknown tags to be found in the tagging task. In general, a retrieval based tagging approach consists of two steps: (1) retrieving a set of visually similar social photos, which are closest to the query photo; and (2) annotating the query photo by a set of most relevant tags that are associated with the retrieved similar photos.

For the first step, there are two typical approaches to find a set of nearest neighbors with respect to a query image. One is to retrieve the k-nearest neighbors of the query image, i.e.,

$$\mathcal{N}_k(x_q) = \{i \in [1, \ldots, n] | x_i \in \text{kNN list}(x_q)\}\ , \tag{14}$$

where n is the total number of photos in the social photo repository. The other way is to retrieve a set of nearest photos within certain distance range, i.e.,

$$\mathcal{N}_\epsilon(x_q) = \{i \in [1, \ldots, n] |\ \|x_i - x_q\|_M \leq \epsilon\}\ , \tag{15}$$

where ϵ is a predefined distance threshold. For both approaches, it is clear that an effective distance metric M is essential to retrieve the set of nearest neighbors. In this paper, we adopt the first approach and employ the metric learned by pRCA to compute the k-NN list.

For the second step, we suggest a simple tag ranking scheme by slightly adapting the idea of majority voting. Specifically, we define a set of candidate tags $\mathcal{T}_w$ as:

$$\mathcal{T}_w = \bigcup_{i\in\mathcal{N}_k} \mathcal{T}_i \tag{16}$$

where $\mathcal{T}_i$ represents the set of tags associated with image I_i. For each candidate tag $w \in \mathcal{T}_w$, we compute its frequency appearing in the k nearest web photos, denoted by $f(w)$. We will then incrementally add the best tag w^* into the tag set of the query image $\mathcal{T}_q = \mathcal{T}_q \cup \{w^*\}$, where

$$w^* = \operatorname*{arg\,max}_{w\in\mathcal{T}_w \wedge w\notin\mathcal{T}_q} \frac{f(w)}{avg\_d(x_q, w) + \kappa} \tag{17}$$

where $avg\_d(x_q, w)$ represents the average distance between the query image and those candidate photos that contain tag w, and κ is a smoothing parameter which is simply fixed to 1 in our experiments. The above formula indicates that we prefer to assign the query image with a tag of *high* frequency and *small* average distance.

7. EXPERIMENTS

This section presents our experimental results on automated photo tagging tasks.

7.1 Experimental Testbed

We collected a large social photo testbed with 205,442 photos crawled from Flickr, in which most photos contain user-tags and other metadata. We split the whole dataset into three disjoint partitions: a *training* set, a *test* set, and a *database* set. We describe the details of the three partitions, respectively.

The training set is used for semantic metric learning. In particular, we randomly sampled 16,588 photos associated with tags from the whole photo testbed. We did not make any refinements on the associated tags. To provide visual words for training the graphical models, we construct the bag-of-visual-words representation by extracting local features from the training photos using SIFT descriptor [19].

The test set is used for evaluating the photo tagging performance. In particular, we randomly picked 2,000 photos from the whole photo testbed as the query images to test the photo tagging performance. To improve the quality of test data, we created the annotation ground truth by manually removing some clear noises to refine the original tags. Since the retrieval by local feature is too time consuming and impractical for large scale dataset, we only adopt the simple global feature for retrieval and annotation experiments.

Finally, the rest social photos in the testbed are engaged as the database set, which serves the base of social photo repository for tagging. Finally, for the photos in both test and database sets, we extract a set of effective and compact visual features, including: (1) grid color moments, (2) edge direction histogram, (3) Gabor textual features, and (4) Local binary pattern histograms. In total, a 297-dimensional feature vector is used to represent each photo. All experiments were run on a PC with 2.8GHz CPU with Matlab.

7.2 Compared Schemes

To examine the effectiveness of our technique, we compare the proposed pRCA algorithm with some baseline and a number of state-of-the-art DML methods, including (1) a **baseline** that simply adopts Euclidean distance, regular RCA [1], Discriminative Component Analysis (DCA) [15], Information-Theoretic Metric Learning (ITML) [6], Large Margin Nearest Neighbor (LMNN) [30], Neighbourhood Components Analysis (NCA) [16], and Regularized Distance Metric Learning (RDML) [23]. Note that we excluded other DML methods in our comparison mainly due to their computational infeasibility for such large-scale applications. For example, the well-known DML method in [32] is only applicable to a very small dataset.

Since no explicit side information is available for traditional DML, in training stage, we performed clustering on training photos using both visual features and tag co-occurrence information. Photos that have similar visual contents and share common tags will be grouped together. Finally, we generate side information from the resulting clusters (after removing trivial clusters) as the inputs for DML.

7.3 Experimental Setup and Protocols

Regarding parameter settings, for the pRCA learning, we assume there are m ($m = 500$) latent chunklets for the N ($N = 16,588$) training examples, and generate an $m \times N$ matrix of probabilistic latent chunklets distribution by the graphical model as the probabilistic side information, which is used as the prior probability matrix P_0 for metric learning. For the extraction of visual words in LCE, we set the number of visual words $A = 1,000$, and the number of tags $W = 2,000$. The parameter γ of pRCA was simply fixed to 0.5 for all experiments. For other DML methods, we adopt the same settings, i.e., 500 chunklets for producing the side information. For their parameters, we chosen them according to the suggestions/empirical results in the original work.

To evaluate the automated photo tagging performance by different methods, we employ the proposed retrieval-based annotation solution presented in Section 5. Firstly, for each query photo in the test set, top k ($k = 30$) nearest photos from the database are first retrieved as the set of candidate images. Then, we annotate the query photo by assigning top t ($t = 1, \cdots, 10$) tags ranked by the function in (17). Finally, we adopt standard average precision and average recall at top t tags as performance metrics to evaluate the automated photo tagging performance.

7.4 Experiment I: Numerical Evaluation

Figure 3 and Figure 4 show average precision and average recall at top t annotated tags, respectively. For these results, we fixed the number of nearest neighbors k to 30 for all compared methods. In both figures, the horizontal axis denotes the number of top tags t that ranges from 1 to 10.

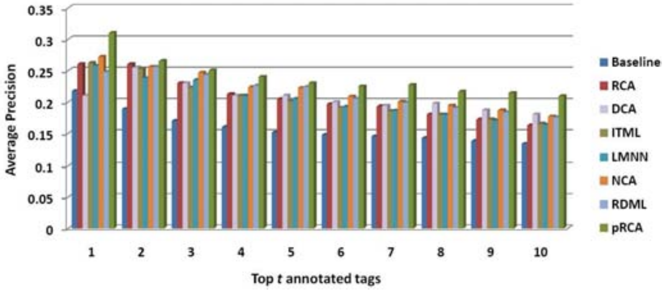


Figure 3: Average precision at top t annotated tags

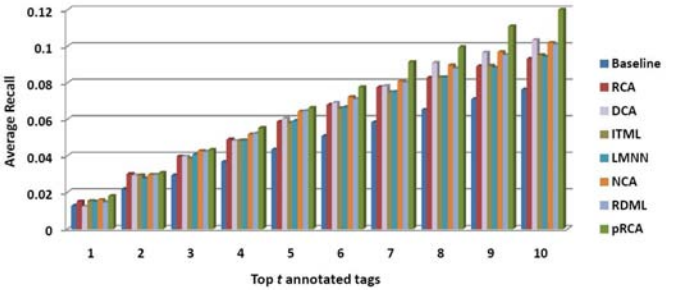


Figure 4: Average recall at top t annotated tags

From the figures, we can draw several observations. First of all, we found that most DML techniques outperformed the baseline by simple Euclidean distance. This shows that DML techniques are beneficial and critical to the retrieval-based photo tagging tasks. Second, we found that for some cases, some DML methods did not perform well, which could be even worse than the Euclidean method. For example, for the case of top-1 annotated tag, we found that DCA performed slightly worse than Euclidean. We believe this is mainly due to the noisy side information issue. This again shows that it is important to develop some effective and robust method in our problem. Further, we observe that the proposed pRCA algorithm considerably outperformed other approaches in most cases. For instance, for the case of top-1 tag, pRCA achieved average precision of about 31%, which improves the baseline approach over 40% and over RCA about 20%. Finally, Figure 5 shows precision-recall curves. Similar observations were found. These results again validate the efficacy and significancy of our technique.

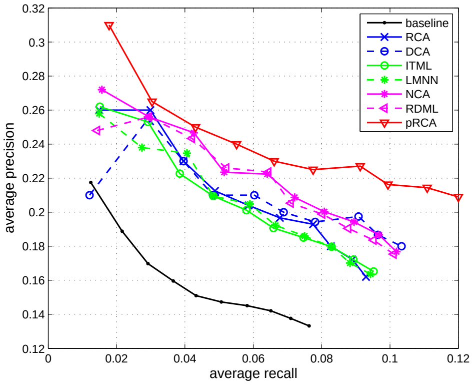


Figure 5: Comparisons of the precision-recall curves

7.5 Experiment II: Evaluation of Varied k

We also notice that the parameter of the number of nearest neighbors k can influence the annotation performance. To evaluate its impact, we examine the annotation performance by varying the value of k. Figure 6 shows the average precision results of the proposed pRCA annotation approach by varying the value of k from 10 to 50.

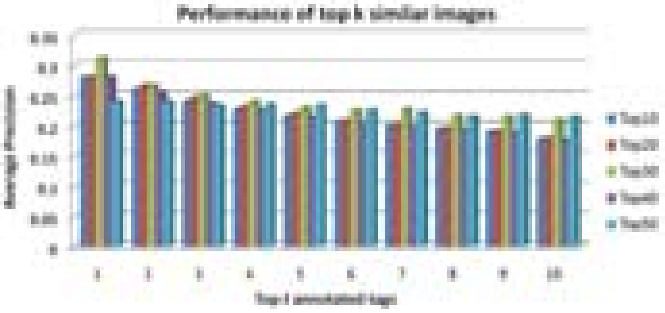


Figure 6: Average precision at top t tags using top k retrieved images by pRCA for annotation.

From the results, we found that when k equals to 30, the resulting performance is generally better than others. In fact, if k is too large, e.g. 50, lots of noisy tags may be included as there may not exist many relevant images in the database. However, if k is too small, some relevant tags may not appear, which again may degrade the performance.

7.6 Experiment III: Time Cost Evaluation

The third experiment is to evaluate the time efficiency performance of the proposed DML algorithm. To this purpose, we compare time performance of our algorithm with other DML algorithms. Table 1 summarizes the time performance evaluation results.

The results showed that the most efficient method is the regular RCA approach, and the worst one is NCA, which is significantly slower than others. Finally, by comparing with other competing algorithms, we found that pRCA is quite competitive, which is worse than RCA,DCA, and RDML, but is considerably better than ITML, LMNN, and NCA.

Table 1: Time cost of different DML methods.

(s)	baseline	RCA	DCA	ITML
Time	N/A	731.63	865.58	1185.27
(s)	LMNN	NCA	RDML	pRCA
Time	1673.23	28989.78	824.81	891.15

7.7 Experiment IV: Qualitative Comparison

The last experiment is to examine qualitative performance of the photo tagging solution. We randomly picked 6 query photos from the test set and showed the qualitative annotation results in Figure 7. From these results, we observe that our solution generally achieves better qualitative results than others.

8. DISCUSSIONS AND LIMITATIONS

Despite encouraging results obtained, our scheme has not yet solved all challenges thoroughly. In particular, we should address several limitations of our work.

Firstly, our method aims to learn a global distance metric for retrieval and annotation. Although global metric is more efficient and scalable for large applications, for some situations, learning a local metric [34] may be more effective. Future work will investigate more effective DML techniques.

Secondly, for efficiency consideration, we only extract global features to represent images in the test and database sets. Global features are usually more effective for some scene annotation tasks, while local features may be more effective for some object annotation tasks [19]. Future work should study the combination of both global and local features.

Thirdly, for the proposed pRCA scheme, the current efficient solution only finds local optima. Although promising results have been achieved by the current solution, we will examine the feasibility of finding global optima.

Finally, the current retrieval-based tagging scheme is generally k-nearest neighbor (k-NN) learning. While k-NN is good for efficiency, it does have some limitations, e.g., linear and no explicit classification model. Future work can study other machine learning techniques, such as kernel methods, to improve photo tagging performance.

9. CONCLUSIONS

This paper addressed a new challenging research problem, i.e., probabilistic distance metric learning (PDML) from uncertain side information that implicitly exists in some real applications. Unlike conventional DML techniques that work with explicit side information, PDML is more challenging given that the side information is explicitly provided. In this paper, we propose a two-step PDML framework, by firstly discovering probabilistic side information from the data using a graphical model approach, and then present an effective probabilistic RCA algorithm to find an optimal metric from the probabilistic side information. We applied the proposed technique for automated photo tagging applications on a social photo testbest of over 200,000 photos from Flickr, and extensively compared our technique with a number of state-of-the-art DML techniques. Encouraging results showed that our technique is effective and promising.

Acknowledgments

The work was supported in part by Singapore MOE Academic Tier-1 Grant (RG67/07) and NRF IDM Grant (IDM-004-018), National Science Foundation (IIS-0643494), and US Army Research Office (ARO Award W911NF-08-1-0403), the National High Technology Research and Development Program of China (863) (2008AA01Z117), the Research Fund for the Doctoral Program of Higher Education (20070358040), and National Natural Science Foundation of China (60672056). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of MOE, NRF and NSF.

10. REFERENCES

[1] A. Bar-Hillel, T. Hertz, N. Shental, and D. Weinshall. Learning a mahalanobis metric from equivalence constraints. *JMLR*, 6:937–965, 2005.

[2] J. C. Bezdek and R. J. Hathaway. Convergence of alternating optimization. *Neural, Parallel Sci. Comput.*, 11(4):351–368, 2003.

[3] G. Carneiro, A. B. Chan, P. Moreno, and N. Vasconcelos. Supervised learning of semantic classes for image annotation and retrieval. *IEEE Tran. PAMI*, pages 394–410, 2006.

[4] G. Carneiro and N. Vasconcelos. Formulating semantic image annotation as a supervised learning problem. In *IEEE CVPR*, pages 163–168, 2005.

[5] W.-Y. Chen, D. Zhang, and E. Y. Chang. Combinational collaborative filtering for personalized community recommendation. In *Proc. 14th ACM SIGKDD Conference*, pages 115–123, 2008.

[6] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon. Information-theoretic metric learning. In *ICML*, pages 209–216, 2007.

[7] P. Duygulu, K. Barnard, J. de Freitas, and D. A. Forsyth. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In *ECCV*, pages 97–112, 2002.

[8] J. Fan, Y. Gao, and H. Luo. Multi-level annotation of natural scenes using dominant image components and semantic concepts. In *ACM Multimedia*, pages 540–547, 2004.

[9] K. Fukunaga. *Introduction to Statistical Pattern Recognition.* Elsevier, 1990.

[10] A. Globerson and S. Roweis. Metric learning by collapsing classes. In *NIPS'05*, 2005.

[11] J. Hayes and A. Efros. Scene completion using millions of photographs. In *SIGGRAPH*, pages 835–846, 2007.

[12] X. He and R. S. Zemel. Learning hybrid models for image annotation with partially labeled data. In *NIPS*, pages 625–632, 2008.

[13] C.-H. Hoi and M. R. Lyu. A novel log-based relevance feedback technique in content-based image retrieval. In *Proceedings of ACM Multimedia Conference (MM2004)*, New York, NY, USA, 2004.

[14] S. C. Hoi, W. Liu, and S.-F. Chang. Semi-supervised distance metric learning for collaborative image retrieval. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR2008)*, June 2008.

[15] S. C. H. Hoi, W. Liu, M. R. Lyu, and W.-Y. Ma. Learning distance metrics with contextual constraints for image retrieval. In *Proc. CVPR2006*, New York, US, June 17–22 2006.

[16] G. H. J. Goldberger, S. Roweis and R. Salakhutdinov. Neighbourhood components analysis. In *NIPS17*, 2005.

[17] J. Jeon, V. Lavrenko, and R. Manmatha. Automatic image annotation and retrieval using cross-media relevance models. In *SIGIR'03*, pages 119–126, Toronto, Canada, 2003.

[18] W. Li and M. Sun. Semi-supervised learning for image annotation based on conditional random fields. In *CIVR*, pages 463–472, 2006.

[19] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60:91–110, 2004.

[20] M. Rege, M. Dong, and J. Hua. Graph theoretical framework for simultaneously integrating visual and textual features for efficient web image clustering. In *WWW '08: Proceeding of the 17th international conference on World Wide Web*, pages 317–326, New York, NY, USA, 2008. ACM.

[21] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman. Labelme: A database and web-based tool for image annotation. *Int. J. Comput. Vision*, 77(1-3):157–173, 2008.

[22] S. Shalev-Shwartz and Y. Singer. Efficient learning of label ranking by soft projections onto polyhedra. *J. Mach. Learn. Res.*, 7:1567–1599, 2006.

[23] L. Si, R. Jin, S. C. H. Hoi, and M. R. Lyu. Collaborative image retrieval via regularized metric learning. *ACM Multimedia Systems Journal*, 12(1):34–44, 2006.

[24] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Trans. PAMI*, 22(12):1349–1380, 2000.

[25] N. Snavely, S. Seitz, and R. Szeliski. Photo tourism: Exploring photo collections in 3d. In *SIGGRAPH*, pages 835–846, 2006.

[26] A. Torralba, Y. Weiss, and R. Fergus. Small codes and large databases of images for object recognition. In *CVPR*, 2008.

[27] C. Wang, L. Zhang, and H.-J. Zhang. Learning to reduce the semantic gap in web image retrieval and annotation. In *SIGIR'08*, pages 355–362, Singapore, 2008.

[28] M. Wang, X. Zhou, and T.-S. Chua. Automatic image annotation via local multi-label classification. In *ACM CIVR*, pages 17–26, New York, NY, USA, 2008. ACM.

[29] X.-J. Wang, L. Zhang, F. Jing, and W.-Y. Ma. Annosearch: Image auto-annotation by search. In *CVPR'06*, pages 1483–1490, 2006.

[30] K. Weinberger, J. Blitzer, and L. Saul. Distance metric learning for large margin nearest neighbor classification. In *NIPS*, 2006.

[31] L. Wu, X.-S. Hua, N. Yu, W.-Y. Ma, and S. Li. Flickr distance. In *Proceeding of 16th ACM international conference on Multimedia (MM'08)*, pages 31–40, Vancouver, British Columbia, Canada, 2008.

[32] E. P. Xing, A. Y. Ng, M. I. Jordan, and S. Russell. Distance metric learning with application to clustering with side-information. In *NIPS2002*, 2002.

[33] R. Yan, A. Natsev, and M. Campbell. A learning-based hybrid tagging and browsing approach for efficient manual image annotation. In *IEEE CVPR'08*, 2008.

[34] L. Yang, R. Jin, R. Sukthankar, and Y. Liu. An efficient algorithm for local distance metric learning. In *AAAI*, 2006.

Appendix A: Proof of Corollary 1

Proof. By fixing μ and P, the optimization reduces to:

$$\min_{M \succeq 0} \quad \sum_{i=1}^{n} \sum_{k=1}^{m} p_i^{(k)} \|x_i - \mu_k\|_M^2 - \lambda \log |M| \tag{18}$$

By differentiating the Lagrangian with respect to M, we have the following equality:

$$\sum_{i=1}^{n} \sum_{j=1}^{k} p_i^{(k)} (\mathbf{x}_i - \mu_k)(\mathbf{x}_i - \mu_k)^\top - \lambda M^{-1} = 0 \tag{19}$$

Hence, we have the optimal solution as: $M = \frac{1}{\lambda} \hat{C}^{-1}$, where the matrix $\hat{C}$ is given as follows:

$$\hat{C} = \sum_{i=1}^{n} \sum_{j=1}^{k} p_i^{(k)} (\mathbf{x}_i - \mu_k)(\mathbf{x}_i - \mu_k)^\top \tag{20}$$

When $p_i^{(k)}$ takes only 0 or 1, it can be seen clearly that the solution of M is almost identical to the solution learned by RCA (up to a global scale factor). Hence, pRCA reduces to regular RCA learning in this special case. □

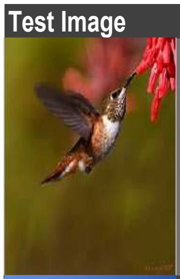
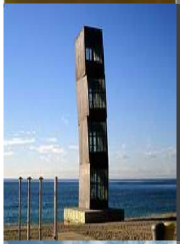
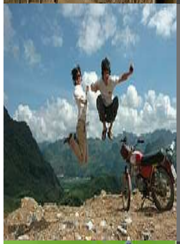
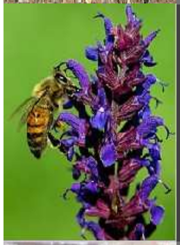
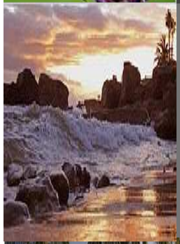
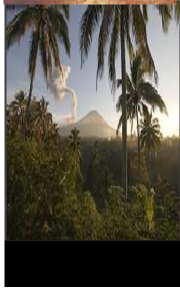
Test Image	Baseline	RCA	DCA	ITML	LMNN	NCA	RDML	pRCA
	**Nature** **Naturesfinest** Abigfave Superaplus Bravo Outstanding Specialanimal **Wildlife** **Birds** Butterfly	**Nature** Anawesome **Naturesfinest** **Specialanimal** **Bird** **Wildlife** Excellence Explore Coast Water	July **Light** **Nature** Butterfly Soe Sculpture **Summer** Wow Blue **Petals**	**Nature** Awesome **NatureFinest** **Specanimal** **Bird** **Wildlife** Excellence Bravo Hiking **Animals**	**Naturefinest** Macro Awesome Abigfave Aplusphoto Super Golden **Nature** San Orange	**Nature** **Red** Awesome Bravo **Green** **Specanimal** Super Magicdonkey Sea Coast	Beach Abigfave Sunset Canon **Green** **Birds** August California Bravo Sky	**Nature** **Bird** **Specialanimal** **Wildlife** **Petals** Bravo **Green** Magicdonkey Soe California
	Abigfave **Sky** **Water** Sunset Anawesome Naturesfinest Australia Mountain Geotagged Roadtrip	**Nature** Sunset **Sky** **Blue** Abigfave **Beach** **Ocean** **Water** **Landscape** **Clouds**	Sunset AbigFave Topf25 **Sky** **Blue** **Landscape** **River** June Church Sign	**Sky** Trees **Blue** Explore **Clouds** Fog Bravo Awesome **Landscape** Himmel	California Hdr Arizona Interesting Explore Car Auto Desert Summer **Water**	Sun **Blue** Sunset **Water** Fog Bravo **Sky** **Nature** Interesting Desert	Abigfave Hdr **Light** **Nature** Aplusphoto **Water** **Clouds** Nikon Photoshop River	**Water** **Sky** Sunset Abigfave Anawesome **Naturesfinest** **Landscape** **Nature** California Sanfrancisco
	Abigfave **Sky** **Clouds** Photomatix Water **Natures finest** Blue Super shot Hdr Masterpiece	Water **Landscape** Hdr Photomatix Blue **Trees** **Clouds** White Birds diamondclass	**Sky** Colorful **Blue** Beautiful Water Super USA Architecture Colors Nikond70	Hdr Photomatix Nikond70 Texas Blended Big Water **Rock** California Nikon	Water Awesome Red **Nature** **Honeymoon** Ireland River Beautiful **Mountain** **Clouds**	**Blue** Water **Rock** **Landscape** Hdr Texas Nikon **Cloud** Bravo River	Sunset **Landscape** **Clouds** **Nature** California Abigfave Bravo Beach **Sky** Water	**Mountains** **Honeymoon** Supershot Masterpiece **Rock** Abigfave **Blue** Hdr Bravo aplusphoto
	Water Diamondclass **Naturesfinest** **Beauty** Red **Specanimal** Anawesome Flickrdiamond Interesting Green	Night Geotagged Sunset Bravo Artlibre Light Nightshot Nikond80 Texas dfw	Orange Awesome Black Building Bravo Colors August **Beauty** Blue Photo	Bravo Magicdonkey Artlibre Blue Hdr **Nature** Perfect Abigfave Water Light	**Nature** 510fav 15fav Amsterdam **Insects** Mountain Snow Male Awesome Supershot	Night Perfect dfw Photo **Bee** **Nature** **Beauty** Colors **Insects** Green	Travel Rock Camping **Spirit** **Nature** California Abigfave Trip Bravo Aplusphoto	**Naturesfinest** **Flower** **Macro** **Nature** **Insect** Nikon **Beauty** Butterfly Florida **specanimal**
	Hdr **Nature** Aplusphoto **Sunset** **Landscape** **Beautiful** **Mountain** Hdri Abigfave bravo	Scotland **Water** **Clouds** Beach **Sunset** Hdr Trip Pond **Sunrise** Newzealand	Hdr **Water** Photomatrix Ireland Sea **Landscape** **Nature** **Sky** **Sunset** Abigfave	Hdr **Landscape** Beach **Clouds** Abigfave **Water** Nikon Canon San Racing	**Nature** **Landscape** Light Home 1on1 Continuum 2for2 **Water** Photos **Rock**	**Sun** **Tree** Light Beach **Water** Canon **Sunrise** Abigfave Photos **Nature**	**Nature** Abigfave **Water** Interesting Explore Trees **Sunset** **Landscape** **Clouds** Impressed	**Cloud** **Nature** **Sunset** **Mountain** **Landscape** Photomatix **Adventure** D200 Hdr Beach
	**Trees** York AbigFave Bravo Landscape Awesome **Nature** **Sunrise** 2006 New	Canon Intesesting **Rock** Beach Sea Red Kids Children Lomo Lomography	Architechture Boston Building **Trees** Amsterdan Cannon Japan Train Graffiti Anaheim	**Fog** Landscape 2006 Old Abandoned Decay Abstract Autumn Fall **Mist**	Hdr Texas Awesome **Mountains** California Photomatix Photoshop **Nature** Landscape Flickrexplore	**Rock** Awesome Red **Trees** Kids Sea **Sky** **Nature** **Fog** Fall	Abigfave Nikon D50 Impressed Super Interesting **Explore** White Gallery UK	**Trees** **Mountain** **Rock** **Cloud** **Sunshine** Beach **Leaves** **Sky** Landscape **Nature**

Figure App2. Top10 Annotation Results

Figure 7: Examples showing the tagging results by eight different methods. For each row, the first image is a test image and each following block shows top 10 tags annotated by one method. The correct tags are highlighted by yellow color.