REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
07-07-2008	Final Report	1-Oct-2007 - 30-Jun-2008

4. TITLE AND SUBTITLE
Final Report on "Content Based Image Retrieval: Application to Tattoo Images for Victim and Suspect Identification"

5a. CONTRACT NUMBER
W911NF-07-1-0665

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611102

6. AUTHORS
Anil K. Jain

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
Michigan State University
Contract & Grant Admin.
Michigan State University
East Lansing, MI 48824 -

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
53241-CI-II.1

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
We have designed a prototype Content-based image retrieval (CBIR) system, called Tattoo-ID, for tattoo image matching and retrieval. CBIR systems automatically determine the image content in the form of low-level image features to compute the similarity between two images, rather than relying on human-assigned (external) class labels. We have examined several key design issues related to building a prototype CBIR system for matching and retrieving tattoo images. Tattoos are imprints on the skin that are being increasingly used by forensics and law enforcement agencies for identifying victims and suspects. Our prototype and first of a kind system demonstrates that it is possible to apply CBIR to this application domain. Initial retrieval

15. SUBJECT TERMS
Image retrieval, forensics, tattoo images, biometrics, CBIR

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Anil Jain
U	U	U	SAR		19b. TELEPHONE NUMBER 517-355-9282

Report Title

Final Report on "Content Based Image Retrieval: Application to Tattoo Images for Victim and Suspect Identification"

ABSTRACT

We have designed a prototype Content-based image retrieval (CBIR) system, called Tattoo-ID, for tattoo image matching and retrieval. CBIR systems automatically determine the image content in the form of low-level image features to compute the similarity between two images, rather than relying on human-assigned (external) class labels. We have examined several key design issues related to building a prototype CBIR system for matching and retrieving tattoo images. Tattoos are imprints on the skin that are being increasingly used by forensics and law enforcement agencies for identifying victims and suspects. Our prototype and first of a kind system demonstrates that it is possible to apply CBIR to this application domain. Initial retrieval results show great promise and pave the way for continued large scale development in this area.

List of papers submitted or published that acknowledge ARO support during this reporting period. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Number of Papers published in peer-reviewed journals: 0.00

(b) Papers published in non-peer-reviewed journals or in conference proceedings (N/A for none)

Number of Papers published in non peer-reviewed journals: 0.00

(c) Presentations

Number of Presentations: 0.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts): 0

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts): 2

(d) Manuscripts

Number of Manuscripts: 0.00

Number of Inventions:

Graduate Students

NAME	PERCENT_SUPPORTED
Jung-Eun Lee	0.50
FTE Equivalent:	**0.50**
Total Number:	**1**

Names of Post Doctorates

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Names of Faculty Supported

NAME	PERCENT_SUPPORTED	National Academy Member
Anil K. Jain	0.10	No
FTE Equivalent:	**0.10**	
Total Number:	**1**	

Names of Under Graduate students supported

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period:	0.00
The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:.....	0.00
The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:.....	0.00
Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):	0.00
Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:	0.00
The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense	0.00
The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields:	0.00

Names of Personnel receiving masters degrees

NAME
Total Number:

Names of personnel receiving PHDs

NAME
Total Number:

Names of other research staff

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Sub Contractors (DD882)

Inventions (DD882)

Rank-based Distance Metric Learning: An Application to Image Retrieval

Jung-Eun Lee, Rong Jin and Anil K. Jain
Michigan State University
East Lansing, MI 48824, USA
{leejun11,rongjin,jain}@cse.msu.edu

Abstract

We present a novel approach to learn distance metric for information retrieval. Learning distance metric from a number of queries with side information, i.e., relevance judgements, has been studied widely, for example pairwise constraint-based distance metric learning. However, the capacity of existing algorithms is limited, because they usually assume that the distance between two similar objects is smaller than the distance between two dissimilar objects. This assumption may not hold, especially in the case of information retrieval when the input space is heterogeneous. To address this problem explicitly, we propose rank-based distance metric learning. Our approach overcomes the drawback of existing algorithms by comparing the distances only among the relevant and irrelevant objects for a given query. To avoid over-fitting, a regularizer based on the Burg matrix divergence is also introduced. We apply the proposed framework to tattoo image retrieval in forensics and law enforcement application domain. The goal of the application is to retrieve tattoo images from a gallery database that are visually similar to a tattoo found on a suspect or a victim. The experimental results show encouraging results in comparison to the standard approaches for distance metric learning.

1. Introduction

Due to rapid growth in the number of available digital images, content-based image retrieval (CBIR) has been extensively studied over the past decade. Most CBIR systems use low-level image features, such as color, texture, and shape, to represent the visual content. These features are automatically extracted from images to compute the similarity between a query and images in the database [7, 17, 25]. However, the retrieval performances of most CBIR systems do not currently meet user expectations. The major reason for this limited performance is that the low-level features are not able to capture the perceived image similarity observed by humans. Consequently, one of the major challenges in CBIR is how to compensate for the *semantic gap* using the low-level features. Several different similarity functions using low-level features have been proposed and examined [12, 19, 28]. Nevertheless, as Santini et al. argued in [20], the only perceptual similarity that can meaningfully be used is pre-attentive similarity, not semantic similarity.

While most CBIR applications emphasize identifying *semantically similar* images, such as "vacation images", there is increasing interest in retrieving *visually similar* images, such as "different images of the White House". The concept of visual similarity is crucial in many real applications like "tattoo image retrieval" for suspect or victim identification [15] that plays an important role in forensic and law enforcement. Because these applications aim to retrieve different images of the same object (e.g., tattoo), semantic perception does not play a major role in retrieval. This fundamental difference makes it more feasible to retrieve visually similar images based only on the low-level visual features.

The key to measure accurate visual similarity between images is to find appropriate distance metric for the given CBIR task. While most existing studies use a pre-defined distance metric for image similarity measurement, our goal is to learn a distance metric from a number of training samples with side information i.e., relevance judgments. This approach can be cast into a standard distance metric learning problem, in which a distance metric is found to keep the queries close to the relevant objects and far away from the irrelevant ones. Unfortunately, as revealed by our empirical study, this strategy does not work well for information retrieval. This is because most distance metric learning algorithms assume that two similar objects are separated by a smaller distance than two dissimilar objects. This assumption may not hold for information retrieval, especially when some queries are far away from all the objects in the database while others are close to many of the objects in the database. In these cases, the distance from a relevant object to a "far away" query may be larger than a distance between

[1]This research was supported by ARO grant W911NF-07-1-0665 and NSF IUC on Identification Technology Research (CITeR).

an irrelevant object and a "close by" query. We aim to address this problem by a rank-based distance metric learning. It overcomes the shortcoming of the existing algorithms by comparing the distance among the relevant and irrelevant objects of only a given query. A specially designed regularizer based on the Burg matrix divergence [13] is introduced to alleviate the over-fitting problem.

The rest of the paper is organized as follows. Section 2 describes related work and Section 3 presents the rank-based approach for distance metric learning within the context of image retrieval. Tattoo image retrieval for suspect or victim identification is described in Section 4 as the application domain and experimental results are provided in Section 5. Finally, we conclude our work in Section 6.

2. Related Work

Learning distance metric from available side information has attracted much interest in recent studies. The side information is usually cast in the form of pairwise constraints. The *must-link* (or equivalence) constraints are the pairs of "similar" objects, and *cannot-link* (or inequivalence) constraints are the pairs of "dissimilar" objects. The optimal distance metric is found such that the objects in must-link constraints are close to each other while the objects in the cannot-link constraints are well separated. A number of algorithms have been developed for learning distance metric from pairwise constraints, including the convex programming approach [27, 16], local distance metric learning [11, 30], relevance component analysis [4], discriminative component analysis (DCA) [14], support vector machine based approaches [21], neighborhood component analysis [9] and its extension [8], maximum-margin nearest neighbor (LMNN) classifier [26], a boosting approach [31] and Bayesian distance metric learning [29].

Most of the algorithms for distance metric learning assume that the objects in a must-link constraint are separated by a smaller distance compared to the objects in a cannot-link constraint. However, this assumption may not hold if the input space is heterogeneous and the distances between objects vary significantly from one location of the input space to another. As a consequence, it is inappropriate to directly compare the distance of any must-link constraint to the distance of any cannot-link constraint. Our proposed a rank-based approach for distance metric learning overcomes this shortcoming by comparing the distance of a must-link constraint to that of a cannot-link constraint only when they are from the "same location" in the input space or associated with the same query.

It is worth mentioning that in addition to the paradigm of learning distance metric from pairwise constraints, there are other approaches for distance metric learning. For instance, in [21] the authors proposed to learn a distance metric from relative comparison. Although the approach in [21] is similar to the spirit of this work, it differs significantly in both the overall formulation and the regularizers used to avoid over-fitting. In [31, 10] the authors present a framework of distance metric learning based on maximum likelihood estimation.

3. Distance Metric Learning

The standard distance metric learning involves pairs of objects that are randomly sampled from a database. On the other hand, in CBIR the pairwise constraints are generated by issuing queries against a given database of images, and visually identifying images from the top retrieved ones that are similar to the query images.

Let $\mathcal{D} = \{x_i, i = 1, \ldots, N_D\}$ denote the collection of images to be retrieved where $x_i \in \mathbb{R}^d$ is a feature vector of size d and represents the ith image. Let $\mathcal{Q} = \{q_i, i = 1, \ldots, N_Q\}$ denote the set of queries that are used to generate the pairwise constraints for distance metric learning. Similar to the images in $\mathcal{D}$, each query image q_i is represented by a vector of d attributes. For each query q_i, we denote by $\{x_{i_1}, \ldots, x_{i_K}\}$ the top K images that are retrieved from $\mathcal{D}$ by the given distance metric A_0. We denote by $y_{i_j} \in \{-1, +1\}$ the relevance judgment for the j-th retrieved image x_{i_j}: $y_{i_j} = +1$ when the retrieved image x_{i_j} is visually similar to the query image q_i, and -1 otherwise. Using the language of pairwise constraints, image x_{i_j} and query q_i form a must-link constraint when $y_{i_j} = 1$ and a cannot-link constraint when $y_{i_j} = -1$. Our goal is to learn a distance metric $A \in \mathbb{R}^{d \times d}$ from the generated pairwise constraints that improves over the existing metric A_0.

3.1. Constraint-based Distance Metric Learning

Before presenting the rank-based approach for distance metric learning, we first present a "typical" distance metric learning approach for image retrieval. The approach exploits the assumption that the distance between images in a must-link constraint tends to be smaller than that for a cannot-link constraint. We refer to this typical approach as "*constraint-based*" for distance metric learning to distinguish it from the proposed "*rank-based*" approach.

Following the framework in [27], the optimal distance metric is learned by minimizing the overall distance of the must-link constraints provided that the images in the cannot-link constraints are well separated. This principle can be cast into the following optimization problem:

$$\begin{aligned} \min_{A \in \mathbb{R}^{d \times d}} \quad & \sum_{i=1}^{N_Q} \sum_{j=1}^{K} \delta(y_{i_j}, +1) d(q_i, x_{i_j}; A) + \frac{\lambda}{2} \mathrm{tr}(AA^T) \\ \text{s. t.} \quad & d(q_i, x_{i_j}; A) \geq 1, \quad \forall y_{i_j} = -1 \\ & A \succeq 0, \end{aligned} \tag{1}$$

where $\delta(y, a)$ is a Dirac delta function that outputs 1 when

$y = a$ and zero otherwise. $d(x, x'; A)$ measures the distance between images x and x' based on the metric A, and is defined as

$$d(x, x'; A) \equiv (x - x')^T A (x - x'). \quad (2)$$

There are two sets of constraints used in the above optimization problem. The first set of constraints, $d(q_i, x_{i_j}; A) \geq 1, \forall y_{i_j} = -1$, ensures the pairs of images in the cannot-link constraints are well separated. The second constraint, $A \succeq 0$, ensures that matrix A is indeed a metric. The objective function in (1) consists of two terms. The first term, i.e., $\sum_{i=1}^{N_Q} \sum_{j=1}^{K} \delta(y_{i_j}, +1) d(q_i, x_{i_j}; A)$, measures the sum of the distance over all the must-link constraints. By minimizing this term, we enforce the images in the must-link constraints to be close to each other. The second term in the objective function, i.e., $\lambda \text{tr}(AA^T)/2$, is introduced to regularize the optimal solution for metric A to be a sparse matrix. This is similar to the quadratic regularizer used in support vector machine (SVM) [5]. Finally, the above problem is a Semi-Definite Programming (SDP) problem and can in general be solved by an interior point method [24].

The main shortcoming of the constraint-based approach is that the distance between objects in must-link constraints may vary significantly from one query to another. As a result, the sum of the distance for all the must-link constraints may be dominated by a small number of queries that are indeed very far from the images in the database $\mathcal{D}$, and most of the optimization effort is spent on reducing the distance for these far away queries. According to the representer theorem, the optimal solution A^* to the optimization problem in (1) can be written as:

$$\begin{aligned} A^* &= \sum_{i=1}^{N_D} \sum_{j=1}^{K} \theta_{i_j} \delta(y_{i_j}, -1)(q_i - x_{i_j})(q_i - x_{i_j})^T \\ &\quad + \eta \sum_{i=1}^{N_D} \sum_{j=1}^{K} \delta(y_{i_j}, +1)(q_i - x_{i_j})(q_i - x_{i_j})^T \quad (3) \end{aligned}$$

where θ_{i_j} and η are weights assigned to each pairwise constraint. As indicated by the above theorem, every must-link constraint (i.e., $y_{i_j} = +1$) is assigned the same weight η. As a consequence, the optimal metric A^* may be dominated by the far away queries.

One may consider improving the above approach by viewing the problem of distance metric learning as a binary classification problem, and cast it into the following optimization problem:

$$\begin{aligned} \min_{A \in \mathbb{R}^{d \times d}} \quad & \sum_{i=1}^{N_Q} \sum_{j=1}^{K} \delta(y_{i_j}, +1) \varepsilon_{i_j} + \frac{\lambda}{2} \text{tr}(AA^T) \\ \text{s. t.} \quad & d(q_i, x_{i_j}; A) \geq 1, \quad \forall y_{i_j} = -1 \\ & d(q_i, x_{i_j}; A) < 1 + \varepsilon_{i_j}, \varepsilon_{i_j} \geq 0, \quad \forall y_{i_j} = +1 \\ & A \succeq 0, \quad (4) \end{aligned}$$

where slack variables $\varepsilon_{i_j} \geq 0$ are introduced to account for the errors in classifying images to be similar. Similar to the previous analysis, we have a representer theorem for the optimal solution A^*, i.e.,

$$A^* = \sum_{i=1}^{N_D} \sum_{j=1}^{K} \theta_{i_j} (q_i - x_{i_j})(q_i - x_{i_j})^T. \quad (5)$$

Note that the weights assigned to must-link constraints by the above optimization problem are no longer a single parameter as in (3). However, the following theorem illustrates that the formulation in (4) indeed puts more emphasis on the distances associated with the far away queries.

Theorem 1 *The problem in (4) is equivalent to the following optimization problem:*

$$\begin{aligned} \min_{A \in \mathbb{R}^{d \times d}} \quad & \sum_{i=1}^{N_Q} \sum_{j=1}^{K} \delta(y_{i_j}, +1) l(d(q_i, x_{i_j}; A)) + \frac{\lambda}{2} tr(AA^T) \\ \textit{s. t.} \quad & d(q_i, x_{i_j}; A) \geq 1, \quad \forall y_{i_j} = -1 \\ & A \succeq 0, \quad (6) \end{aligned}$$

where $l(d) = \max(0, d - 1)$.

The above result follows the fact $\xi_{i_j} = \max(0, d(q_i, x_{i_j}; A) - 1)$. As indicated by the above theorem, the loss function $l(d)$ removes any must-link constraint whose distance d is less than 1, and as a result, the impact of far away queries is further amplified by $l(d)$.

3.2. Rank-based Distance Metric Learning

To address the problem when the input space is heterogeneous and the distance in must-link constraints may vary significantly from one query to another, we propose to learn the distance metric learning by a rank-based approach. In particular, instead of requiring the distance of any must-link constraint to be smaller than that of a cannot-link constraint, we only compare the distances of pairwise constraints that are generated by the same query. Hence, a must-link constraint is supposed to have a smaller distance than a cannot-link constraint only when they are from the same query. We cast this idea into the following optimization problem:

$$\begin{aligned} \min_{A \in \mathbb{R}^{d \times d}} \quad & \sum_{i=1}^{N_Q} \sum_{k,j=1}^{K} \delta(y_{i_j}, -1) \delta(y_{i_k}, +1) \varepsilon_{j,k}^i + \frac{\lambda}{2} \text{tr}(AA^T) \\ \text{s. t.} \quad & d(q_i, x_{i_j}; A) - d(q_i, x_{i_k}; A) \geq 1 - \varepsilon_{j,k}^i, \varepsilon_{j,k}^i \geq 0 \\ & A \succeq 0 \quad (7) \end{aligned}$$

Note that a slack variable $\varepsilon_{j,k}^i \geq 0$ is introduced when comparing a must-link constraint (i.e., $y_{i_k} = +1$) and a cannot-link constraint (i.e., $y_{i_k} = -1$) that share the same query. Since only the constraints sharing the same query will be

compared in computing the distance metric, we only require the distance of a must-link constraint to be *relatively* small compared to the distance of a cannot-link constraint and therefore avoid the shortcoming of the constraint-based approach for distance metric learning.

Although the formulation in (7) addresses the shortcomings of the constraint-based approach, it does not take into account the existing distance metric A_0 when learning a new distance metric from pairwise constraints. This could be important if A_0 is engineered by the domain expert to take into account the domain knowledge. It will also be useful to take into account A_0 if we learn the distance metric A in a sequential manner and A_0 is a distance metric learned from the pairwise constraints collected in the previous iterations. In order to explicitly take into account A_0, we replace the regularizer $\lambda \text{tr}(AA^T)/2$ with the Burg matrix divergence [13] that is defined as follows:

$$D(A, A_0) = \text{tr}(AA_0^{-1}(AA_0^{-1})^T) - 2\log\det(AA_0^{-1}) - d. \quad (8)$$

Since A and A_0 may share a different scaling, we normalize matrix A_0 as follows before computing the divergence,

$$\widehat{A_0} = A_0 \frac{\text{tr}(A)}{\text{tr}(A_0)}.$$

Using the above matrix divergence, the problem in (7) is modified as follows:

$$\begin{aligned} \min_{A\in\mathbb{R}^{d\times d}} \quad & \sum_{i=1}^{N_Q}\sum_{k,j=1}^{K} \delta(y_{i_j}, -1)\delta(y_{i_k}, +1)\varepsilon_{j,k}^i + \frac{\lambda}{2} D(A, \widehat{A_0}) \\ \text{s. t.} \quad & d(q_i, x_{i_j}; A) - d(q_i, x_{i_k}; A) \geq 1 - \varepsilon_{j,k}^k, \varepsilon_{j,k}^i \geq 0 \\ & A \succeq 0. \end{aligned} \quad (9)$$

By minimizing the divergence between A and A_0, we require the learned distance matrix A to be similar to A_0.

Remark To better understand the matrix divergence $D(A, A_0)$ in (8), we consider the special case when both A and A_0 are diagonal matrices, i.e., $A = \text{diag}(a_1, \ldots, a_d)$ and $A_0 = \text{diag}(b_1, \ldots, b_d)$. The divergence is now simplified as follows:

$$\begin{aligned} D(A, \widehat{A_0}) &= \sum_{i=1}^{d} \left(a_i/\widehat{b}_i\right)^2 - 2\sum_{i=1}^{d} \log(a_i/\widehat{b}_i) - d \\ &\approx \sum_{i=1}^{d} (a_i/\widehat{b}_i - 1)^2, \end{aligned}$$

where $\widehat{b}_i = b_i \sum_{i=1}^d a_i / (\sum_{i=1}^d b_i)$. The above approximation follows the inequality $\log x \approx x - 1$. When A_0 is an Identity matrix, the divergence $D(A, \widehat{A_0})$ is further approximated as

$$D(A, \widehat{A_0}) \approx \frac{1}{\overline{a}^2} \sum_{i=1}^{d} (a_i - \overline{a})^2,$$

where $\overline{a} = \sum_{i=1}^d a_i/d$. The above analysis indicates that when A_0 is an Identity matrix, the matrix divergence $D(A, \widehat{A_0})$ essentially measures the variance in the diagonal elements of matrix A. Thus, by minimizing the divergence, the resulting matrix A tends to have a flat distribution over its diagonal elements.

3.3. Efficient Implementation

The distance metric learning algorithm described above requires finding the optimal matrix A. This is usually computationally expensive because (i) the number of elements in A is quadratic in the number of dimensions used to represent images, and (ii) the requirement that A has to be positive semi-definite. We reduce the computational cost by assuming A to be a diagonal matrix, i.e., $A = \text{diag}(a_1, \ldots, a_d)$, such that $d(x, x'; A) = d(x, x'; a) = \sum_{i=1}^d (x_i - x_i')^2 a_i$. Then, the problems in (1) and (7) are simplified as

$$\begin{aligned} \min_{a\in\mathbb{R}^d} \quad & \sum_{i=1}^{N_Q}\sum_{j=1}^{K} \delta(y_{i_j}, +1) d(q_i, x_{i_j}; a) + \frac{\lambda}{2}\sum_{i=1}^{d} a_i^2 \\ \text{s. t.} \quad & d(q_i, x_{i_j}; A) \geq 1, \quad \forall y_{i_j} = -1 \\ & a_i \geq 0, i = 1, \ldots, d \end{aligned} \quad (10)$$

and

$$\begin{aligned} \min_{a\in\mathbb{R}^d} \quad & \sum_{i=1}^{N_Q}\sum_{k,j=1}^{K} \delta(y_{i_j}, -1)\delta(y_{i_k}, +1)\varepsilon_{j,k}^i + \frac{\lambda}{2}\sum_{i=1}^{d} (a_i - \overline{a})^2 \\ \text{s. t.} \quad & d(q_i, x_{i_j}; a) - d(q_i, x_{i_k}; A) \geq 1 - \varepsilon_{j,k}^k, \varepsilon_{j,k}^i \geq 0 \\ & a_i \geq 0, i = 1, \ldots, d, \end{aligned} \quad (11)$$

respectively. In the above, an Identity matrix is assumed for A_0. Both problems in (10) and (11) can be solved by standard quadratic programming techniques.

Remark It is interesting to examine the regularizer $\sum_{i=1}^d (a_i - \overline{a})^2$ from the view point of Laplacian. We can rewrite the regularizer into the matrix form, i.e.,

$$\sum_{i=1}^{d} (a_i - \overline{a})^2 = a^\top (I - \mathbf{1}\mathbf{1}^\top/n) a = a^\top L a$$

where L is indeed a graph Laplacian constructed from a fully connected graph with every edge weighted equally. If we have more knowledge regarding the features, we can adopt a different weight for the pairwise relationship between any two features, which will lead to a very different graph Laplacain.

Figure 1. Examples of tattoos belonging to well known gangs: (a) Brazers, (b) Latin Kings, (c) Family Stones, and (d) Insane Deuces [1]

Figure 2. Illustration of large intra-class variability in tattoo images. All the above images belong to the FIRE category

4. Tattoo Images for Victim and Suspect Identification

Tattoos engraved on human body are routinely used to assist in human identification in forensics applications. This is not only because of the increasing prevalence of tattoos, but also due to their impact on other methods of human identification such as visual, pathological, or trauma-based identification [22]. The role of tattoos is particularly important when the primary biometric traits, e.g., fingerprints or face, are either no longer available, or corrupted (e.g. victims of Asian Tsunami and 9/11 terrorist attack). A study by Burma [6] found that delinquents are significantly more likely to have tattoos than non-delinquents which indicates that tattoos could provide a source of information for determining gang membership. Many law enforcement agencies maintain a database of tattoos, i.e., tattoo field in the Computerized Criminal History Records, and it is now a common practice to photograph and catalog tattoo patterns to identify victims and criminals (e.g., gang membership, see Figure 1) [23, 1]. While a tattoo does not uniquely establish the identity of a suspect or a victim, it helps in narrowing down the possible identities since tattood often indicate gang membership, religious beliefs, previous conviction, military services, etc.

The ANSI/NIST-ITL 1-2000 document [3] contains classification standards for tattoo images. The standard has eight major tattoo classes, such as human, animal, symbol, etc, and 80 subclasses. Current practice in law enforcement agencies is to match a query tattoo by performing manual searches in the tattoo database based on matching the class labels. This process is subjective, has limited performance and is time-consuming. Further, a simple class descriptor of a tattoo textual query does not contain all the semantics in the tattoo images as evident by the large intra-class variability (see Figure 2).

Jain et al. proposed a CBIR system for tattoo image matching and retrieval [15]. Although this system showed promising results, its performance is limited because it employs a predefined similarity measure without appropriately weighting different features. We aim to improve its performance by applying the proposed rank-based distance metric learning framework.

4.1. Tattoo Image Database

We use the same tattoo database as in [15], which contains 2,157 tattoo images downloaded from the web [2] and belonging to eight main classes and 20 subclasses in the ANSI/NIST standard [3]. Multiple acquisition of the same tattoo may look different because of various imaging condition, such as brightness, viewpoint and distance (see Figure 3). A tattoo image retrieval system should be invariant to these imagining conditions. To simulate the various imaging conditions, we follow the work in [15] and generate 20 transformed images for every tattoo image in the database (see Figure 4). This results in a total of 43,140 synthesized images.

4.2. Image Features

We choose the low level image attributes same as in [15], i.e., color, shape and texture. The overall size of the feature vector is 272. Similar features have also been used in many other CBIR systems and summarized below.

Color Two color descriptors, color histogram and color correlogram, are extracted from the RGB space. A color correlogram stores the probability of finding a pixel of color j at a distance k from a pixel of color i in the image. The color histogram and correlogram are calculated by dividing each color component into 20 and 63 bins, resulting in a total of 60 and 189 bins for the color histogram and correlogram, respectively. For computational efficiency, we compute color autocorrelogram only between identical colors in a local neighborhood, i.e., $i = j$ and $k = 1, 3, 5$.

Shape Based on 2nd and 3rd order moments, a set of seven features that are invariant to translation, rotation, and scale are obtained. Two different feature sets are extracted, one from the segmented grayscale and the other from gradient tattoo images.

Texture Edge Direction Coherence Vector stores the ratio of coherent to non-coherent edge pixels with the same quantized direction (within an interval of 10 degree). A threshold (0.1% of image size) on the edge-connected components in a given direction is used to decide the region coherency. This feature discriminates structured edges from randomly distributed edges.

The histogram intersection based approach used in [15] to measure image similarity, is used here as the baseline

Figure 3. Eight different images of a butterfly tattoo taken under different imaging conditions

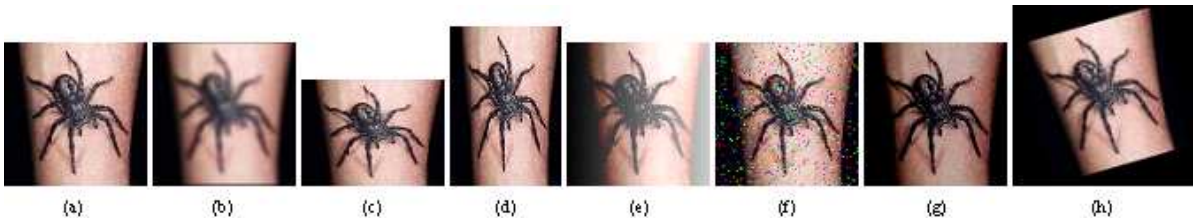

Figure 4. Examples of tattoo image transformation: (a) original, variations due to (b) blurring, (c) and (d) aspect ratio change, (e) illumination, (f) additive noise (g) color transformation, and (h) rotation

performance. This similarity measure calculates the overlapping area between two normalized histograms.

5. Experimental Results

We evaluate the proposed algorithm for distance metric learning on tattoo image retrieval problem. We assume that the query tattoo images are taken under imperfect imaging conditions and therefore can be simulated by the transformed images that were described in Section 4. A retrieved image is deemed to be relevant, when the query image was generated from the retrieved image, by one of the image transformations shown in Figure 4. The number of queries is 43,140 and the size of the database is 2,157. The distance metric is learned off-line from a pool of training examples and, as a result, the matching procedure using the learned distance metric takes the same time as the baseline.

Since there is only one true "similar" image in the database for every query image, we adopt the cumulative matching characteristic (CMC) [18] curve as the evaluation metric. This metric cumulates the correct number of retrieved images as the rank is increased. For cross validation, we divided the database of query images (43,140 images) into ten folds of equal size. One fold of query images is selected for testing, and 5,000 images are randomly selected from the remaining nine folds for training. This procedure is repeated for every fold of query images and the CMC curve, averaged over 10 experiments, is reported with the mean of standard deviations, σ, of all ranks.

Before presenting our results on rank-based distance learning, we will first examine the hypothesis that is used by many other distance metric learning algorithms, namely a distance between two similar object in a must-link pair is usually smaller than the distance between two dissimilar objects in a cannot-link pair. Figure 5 shows the distance distributions based on histogram intersection for both must-link pairs and cannot-link pairs. We notice that the distance distribution for the must-link pairs indeed has a long tail, which makes it difficult to differentiate them from cannot-link pairs. This suggests that the hypothesis assumed by many distance metric learning algorithms may not hold in our image retrieval problem. In this experimental study, we aim to address three important questions:

- Will the rank-based framework be more effective than the constraint-based framework for distance metric learning in the case of image retrieval?
- How important is the regularizer in learning a distance metric for image retrieval?
- How to efficiently train a distance metric by the rank-based framework?

Comparison of Distance Metric Learning Algorithms We now compare the rank-based approach for distance metric learning to the constraint-based approach. Figure 6 shows the retrieval performance of the two distance metric learning approaches. First, we observe that the rank-based approach significantly outperforms the constraint-based approach at every rank. For instance, the rank-1 retrieval accuracy of the ranked-based approach is over 71% while the accuracy of the constraint-based approach is less than 65%. Besides, the constraint-based approach shows very little improvement over the baseline. In fact, it performs noticeably worse than the baseline for the first 5 ranks. This result implies that directly comparing the distance of any must-link constraint to that of any cannot-link constraint may be inappropriate if the input space is heterogeneous. Overall, we observe a significant improvement made by the ranked-based approach for distance metric learning in comparison to the baseline approach, suggesting that the pro-

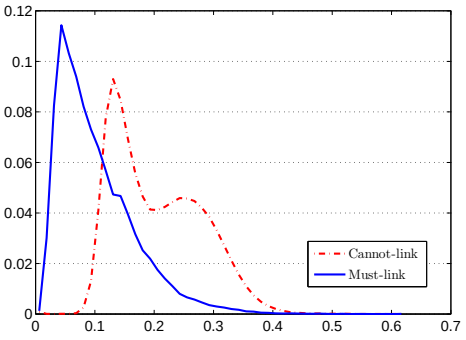

Figure 5. Distance distributions for must-link and cannot-link pairs

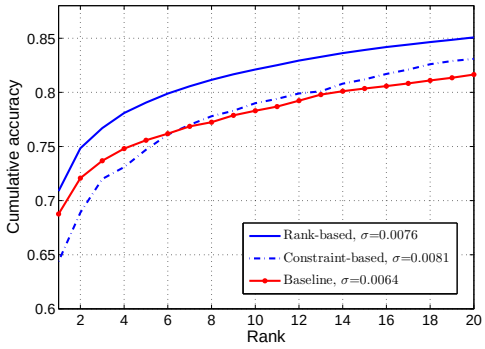

Figure 6. Retrieval accuracy of the rank-based approach and the constraint-based approach for distance metric learning

posed ranked-based approach is effective in handling heterogeneous input space.

Effects of Regularizer This experiment examines the effect of the regularizer in (10) by varying the value of the regularization parameter λ. Figure 7 summarizes the retrieval performance of the rank-based approach with different value of λ. Without a regularizer, i.e., $\lambda = 0$ in (10), the retrieval performance of the rank-based approach is similar to baseline. By increasing the value of regularization parameter from 1 to 100, we observe the overall increase in the retrieval performance. These results indicate the importance of regularizer for distance metric learning. Also, the overall monotonic trend with increasing value of λ makes it relatively easy to choose the appropriate value for λ. In fact, the retrieval performance remains almost unchanged when the regularization parameter passes a certain threshold. We found that the threshold value for the regularization parameter depends the size of training set. In particular, we observed a larger value for the threshold of the parameter when the size of training example is increased.

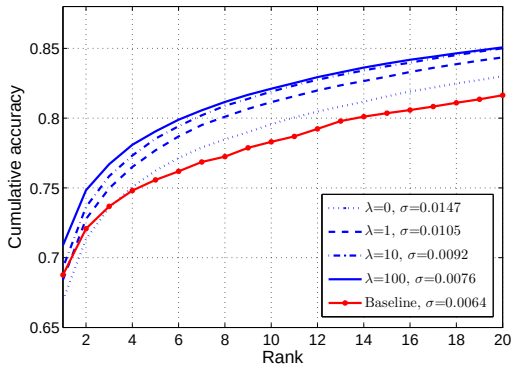

Figure 7. Retrieval accuracy of rank-based approach using different regularization parameter values

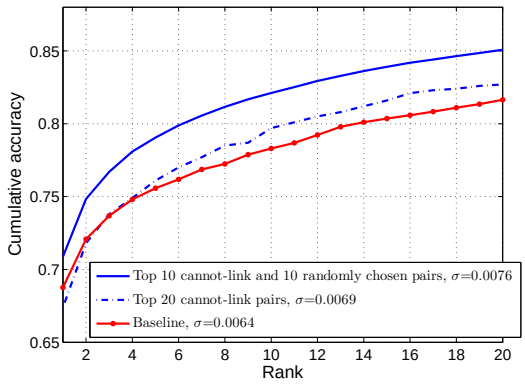

Figure 8. Retrieval accuracy of rank-based approach for distance metric learning using different training pairs

Efficient Training for Distance Metric Learning There are a total of 93 million image pairs in our experiments. It is thus computationally infeasible to use all the pairs for training. Instead, we focus on training the distance metric by selecting "critical" image pairs. The critical image pairs for each query image are formed by the top list of irrelevant images that are retrieved by the baseline approach. In addition, to preserve the diversity of the training pairs, we also randomly select a few images for each query to form additional cannot-links. Figure 8 shows the results of the rank-based approach that is trained by two different sets of pairs: (i) the critical pairs formed by the top ranked 20 irrelevant images, and (ii) the critical pairs formed by the top ranked 10 irrelevant images and 10 randomly selected images. The results show that although the same number of cannot-link sets are used in both the experiments, the distance metric trained from the combination of top ranked images and randomly chosen images performs much better. We attribute the difference to the fact that the top ranked irrelevant images may not be able to represent the feature distribution of images

in the entire database. The randomly chosen images from outside of top ranked images provide general information about the input space while the top rank images supply detailed information only among a given query and irrelevant images.

6. Conclusions

In this paper, we examined the problem of distance metric learning under the context of image retrieval. We presented a rank-based framework for distance metric learning that explicitly addresses the problem of heterogeneous input space. Our approach distinguishes from the previous approach, e.g., pairwise constraint-based distance metric learning, in that it does not assumes shorter distances among relevant objects compared to the distance between objects. The experimental results show that our approach is more effective than the existing algorithms.

References

[1] Gang Ink, http://www.gangink.com/. 5

[2] Online Tattoo Designs, http://www.tattoodesign.com/gallery/. 5

[3] ANSI/NIST-ITL 1-2000 standard: American National Standard for Information Systems - data format for the interchange of fingerprint, facial, & scar mark & tattoo (SMT) information, 2000. 5

[4] A. Bar-Hillel, T. Hertz, N. Shental, and D. Weinshall. Learning distance functions using equivalence relations. In *Proc. ICML*, pages 11–18, 2003. 2

[5] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998. 3

[6] J. H. Burma. Self-tattooing among delinquents: A research note. *Sociology and Social Research*, 43:341–345, 1959. 5

[7] M. Flickner, H. Sawhney, W. Niblack., J. Ashley, Q. Huang, B. Dom, M. Gorkani, J.Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker. Query by image and video content: The QBIC system. *IEEE Computer*, 28:22–32, 1995. 1

[8] A. Globerson and S. Roweis. Metric learning by collapsing classes. In *NIPS*, volume 18, pages 451–458, 2006. 2

[9] J. Goldberger, S. Roweis, G. Hinton, and R. Salakhutdinov. Neighbourhood components analysis. In *NIPS*, volume 17, pages 513–520, 2005. 2

[10] R. Haralick and L. Shapiro. *Computer and Robot Vision II*. Addison-Wesley, 1993. 2

[11] T. Hastie and R. Tibshirani. Discriminant adaptive nearest neighbor classification. *PAMI*, 18(6):607–616, 1996. 2

[12] X. He, W.-Y. Ma, and H.-J. Zhang. Manifold ranking based image retrieval. In *ACM Multimedia*, pages 9–16, 2004. 1

[13] N. J. Higham. Matrix nearness problems and applications. In M. J. C. Gover and S. Barnett, editors, *Applications of Matrix Theory*, pages 1–27. Oxford University Press, 1989. 2, 4

[14] S. C. H. Hoi, W. Liu, M. R. Lyu, and W.-Y. Ma. Learning distance metrics with contextual constraints for image retrieval. In *Proc. CVPR*, pages 2072–2078, 2006. 2

[15] A. K. Jain, J.-E. Lee, and R. Jin. Tattoo-ID: Automatic tattoo image retrieval for suspect & victim identification. In *Proc. Pacific-Rim Conf. on Multimedia*, 2007. 1, 5

[16] J. T. Kwok and I. W. Tsang. Learning with idealized kernels. In *Proc. ICML*, pages 400–407, 2003. 2

[17] W. Ma and B. S. Manjunah. Netra: A toolbox for navigating large image databases. *Multimedia Systems*, 7:184–198, 1999. 1

[18] H. Moon and P. J. Phillips. Computational and performance aspects of pca-based face recognition algorithm. *Perception*, 30:303–321, 2001. 6

[19] H. Muller, T. Pun, and D. Squire. Learning from user behavior in image retrieval: Application of market basket analysis. *Int. J. of Computer Vision*, 56:65–77, 2004. 1

[20] S. Santini and R. Jain. Visual navigation in perceptual database. In *Int. Conf. Visual Information Systems*, pages 101–108, 1997. 1

[21] M. Schultz and T. Joachims. Learning a distance metric from relative comparisons. In *NIPS*, volume 16, pages 41–48, 2004. 2

[22] T. Thompson and S. Black. *Forensic Human Identification, An Introduction*. CRC Press, 2007. 5

[23] B. Valentain. *Gangs and Their Tattoos: Identifying Gangbangers on the Stree and in Prison*. Paladin Press, 2000. 5

[24] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Review*, 38(1):49–95, 1996. 3

[25] J. Z. Wang, J. Li, and G. Wiederhold. SIMPLIcity: Semantics-sensitive intergrated matching for picture libraries. *PAMI*, 23:947–963, 2001. 1

[26] K. Weinberger, J. Blitzer, and L. Saul. Distance metric learning for large margin nearest neighbor classification. In *NIPS*, volume 18, pages 1473–1480, 2006. 2

[27] E. Xing, A. Ng, M. Jordan, and S. Russell. Distance metric learning with application to clustering with side-information. In *NIPS*, volume 15, pages 505–512, 2003. 2

[28] R. Yan, A. Hauptmann, and R. Jin. Negative pseudo-relevance feedback in content-based video retrieval. In *ACM Multimedia*, pages 343–346, 2003. 1

[29] L. Yang, R. Jin, and R. Sukthankar. Bayesian active distance metric learning. In *UAI*, 2007. 2

[30] L. Yang, R. Jin, R. Sukthankar, and Y. Liu. An efficient algorithm for local distance metric learning. In *Proc. AAAI*, pages 450–459, 2006. 2

[31] J. Yu, J. Amores, N. Sebe, and Q. Tian. Toward robust distance metric analysis for similarity estimation. In *Proc. CVPR*, pages 316–322, 2006. 2