INVITED PAPER

Active Learning for Interactive Multimedia Retrieval

Algorithms that employ feedback from users to guide the search process can provide relatively rapid and efficient results from large multimedia data collections.

By Thomas S. Huang, *Fellow IEEE*, Charlie K. Dagli, Shyamsundar Rajaram,
Edward Y. Chang, Michael I. Mandel, Graham E. Poliner, and Daniel P. W. Ellis

ABSTRACT | As the first decade of the 21st century comes to a close, growth in multimedia delivery infrastructure and public demand for applications built on this backbone are converging like never before. The push towards reaching truly ***interactive*** multimedia technologies becomes stronger as our media consumption paradigms continue to change. In this paper, we will profile a technology leading the way in this revolution: ***active learning***. Active learning is a strategy that helps alleviate challenges inherent in multimedia information retrieval through user interaction. We will show how active learning is ideally suited for the multimedia information retrieval problem by giving an overview of the paradigm and component technologies used with special attention given to the application scenarios in which these technologies are useful. Finally, we give insight into the future of this growing field and how it fits into the larger context of multimedia information retrieval.

KEYWORDS | Active learning; content-based information retrieval; human-computer interaction; image/video search; interactive pattern recognition; relevance feedback; user-centered design

I. INTRODUCTION

Interactive multimedia applications are becoming more commonplace as both media delivery infrastructure and the success of new participatory web technologies such as Web 2.0 [1] continue to mature. One consequence of this growth is the emergence of a sophisticated user base more skilled in learning and adapting to new technologies and interaction paradigms than in years past. This user aptitude, but more specifically the potential to monetize this aptitude through focused advertising, has not gone unnoticed by media providers: consider the $580 million purchase of MySpace.com by News Corporation in 2005 [2].

The recent explosion in multimedia-oriented web technologies straddles the intersection between improved media delivery and tech-savvy users. The great popularity of social networking sites such as MySpace [3] and FaceBook [4], image sharing websites like Flickr [5] and Zooomr [6], as well as the gold rush on video sharing sites sparked by Google's $1.65 billion acquisition of YouTube [7] has the potential to usher in a new era in web economics.

Though consumers have quickly embraced multimedia technologies such as image and video search, commercial content-based indexing and interactive search of multimedia documents is still in its nascent stages. With the exception of a beta-version of face-image search functionalities by Google [8] and Exalead [9], respectively, most commercial systems still adapt standard text retrieval technologies to index and search these databases. Finding new and useful interaction paradigms for a tech-aware user base also remains a challenging domain for the research community. It is clear users want to connect and interact, but it is unclear how to leverage this interaction for content organization and search.

Given this commercial playing field, the potential impact of a great advance in interactive multimedia search justifies the continued interest in this well-studied research area. The unique aspect of multimedia information retrieval is that database documents are indexed holistically using *multiple* modalities, by extracting low-level features from each of these modalities such as color, texture, motion, audio timbre, etc. For more focused search applications (such as object detection and retrieval) mid-level features such as shape and trajectory can be used.

Manuscript received May 14, 2007; revised November 29, 2007.

T. S. Huang, **C. K. Dagli**, and **S. Rajaram** are with the Beckman Institute for Advanced Science and Technology, University of Illinois, Urbana, IL 61801 USA (e-mail: huang@ifp.uiuc.edu; dagli@ifp.uiuc.edu; shyamsundar.rajaram@hp.com).
E. Y. Chang is with the Electrical and Computer Engineering Department, University of California, Santa Barbara, CA 93106-9560 USA (e-mail: echang@ece.ucsb.edu).
M. I. Mandel, **G. E. Poliner**, and **D. P. W. Ellis** are with the Department of Electrical and Computer Engineering, Columbia University, New York, NY 10027 USA (e-mail: mim@ee.columbia.edu; graham@ee.columbia.edu; dpwe@ee.columbia.edu).

Digital Object Identifier: 10.1109/JPROC.2008.916364

In order to use low-level information extracted from visual and audio information properly, designers must balance the tradeoff between the extra discriminative ability gained from these modalities with the semantic ambiguity of matching documents based purely on low-level features. This phenomenon is known as the *semantic gap* [10] and motivates the use of *interactive* strategies in designing such search systems.

To further motivate our discussion, consider the following application domains whose technologies seek to bridge the divide between basic research and the potential of the new web economy.

A. Image and Video Search and Retrieval

Commercial image and video search systems have garnered great attention in past years by making multimedia content indexed through text meta-data or filenames searchable. These systems, however, do not index or compute similarity among images or videos using the actual *content* of the documents themselves.

This challenging task is the domain of the well-studied area of content-based image/video retrieval (CBIR) [11], [12]. This field seeks to utilize low-level information seen in the content itself, along with information gained from user interaction to improve search.

Consider, for example, using Google Images to perform a keyword search and then being able to interact with the system by telling it which of the returned images are relevant to the intended query and having it respond to this feedback by adjusting the ranked list of returns accordingly. This interactive paradigm is known as *relevance feedback* [10], [13] and is an approach intended to close the gap between semantic notions of search relevance and the low-level representation of content seen in multimedia documents [14], [15]. This has been the main crux of research in interactive multimedia search technologies over the past 15 years.

Truly interactive content-based systems have the potential to not only improve online image and video search but also many other application areas as well. Progress has been made in the areas of trademark search for digital rights management [16], objectionable content detection [17], sports video analysis and search [18], broadcast news video search [19], as well as other domains which abstract to the content-based retrieval problem. For a broader overview, the reader is referred to [11] and [12].

B. Music Organization and Search

Despite its rocky courtship (and with the help of nearly continuous legal counseling [20]) the tenuous marriage between the recording industry and the online economy continues to grow. Despite constant industry objections, the web is becoming the preferred medium for music delivery. The massive success of online music, in particular Apple's iPod and iTunes [21], serves not only as a cultural and technological landmark but an economic one as well.

Through the development of playlist sharing, as in Apple's iTunes, or interplayer song sharing, as in Microsoft's Zune player [22], it is clear the industry sees music organization and search as a potentially profitable, opportunity-rich area for interactive technologies. Music organization and search presents researchers with wonderfully challenging technical problems.

As music libraries grow, identifying music to listen to or buy becomes increasingly difficult. Music recommendation services such as Pandora [23] provide a pseudo-content-based strategy for interactive playlist generation, but rely on a large collective-labeling project for its aspect profiles. Over recent years, work from the interactive content-based perspective has tried to lift this constraint by envisioning automatic indexing of music files for recommendation and search [24], [25]. In addition, analysis and search of information from the audio track can be used in a general scenario for related or contextual media delivery [26].

II. GENERAL TECHNICAL PARADIGM

For any of these application domains, it is imperative to choose a technical paradigm that best fits the general interactive multimedia search problem. Since we are seeking to represent, model, and search among large collections of data interactively, we must draw on principles from machine learning, data mining, and information retrieval collectively. To illustrate some subtleties of the problem, we ask the reader to consider briefly a typical usage scenario.

Kevin is using an interactive, online image search system to find images for a class project. In particular, he is interested in finding images of his favorite golfer, Tiger Woods. He can either type a text query such as "tiger woods" into the system, or provide to it an image similar to what he is interested in, his *query by example* [10]. The system uses information from the example clip or text query to infer which images in its database most closely resembles his initial search concept. From the ranked list of results, Kevin can try to reformulate the query by interacting with the system. The system readjusts its understanding of the query concept based on information gleaned from this interaction and returns a new set of results. This continues until he is happy with the results. This paradigm is illustrated in Fig. 1.

Designers of such systems must take into consideration the fact that Kevin's time and effort are at a premium. In all search and retrieval applications, the user wants to expend as little effort as possible while being presented with a large number of relevant results quickly. As a result, in order to design successful interactive multimedia retrieval systems, we must balance this constraint across two interrelated system components: the *learning/search strategy* and the *interaction strategy*.

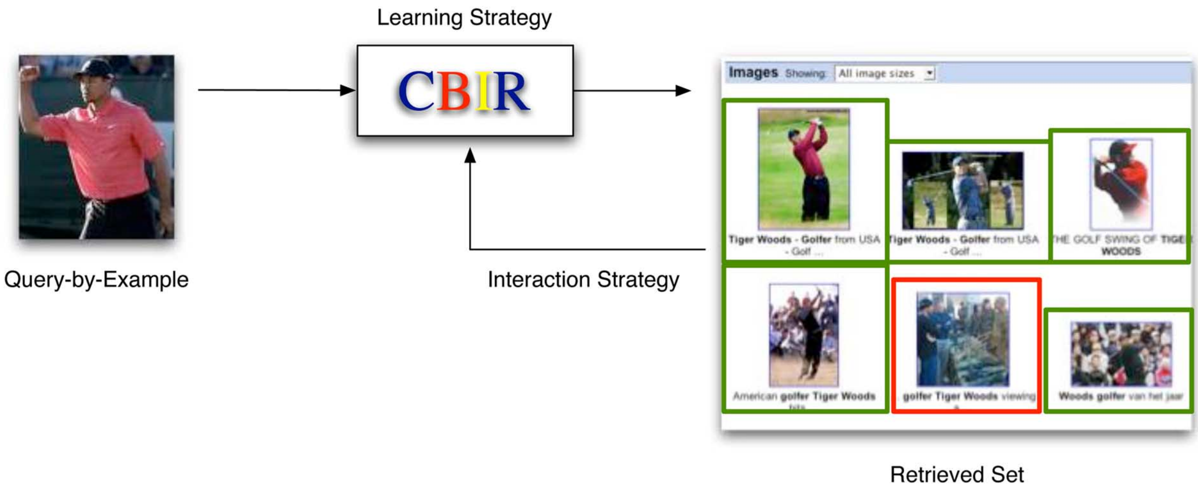

Learning Strategy
CBIR
Query-by-Example
Interaction Strategy
Images Showing: All image sizes
Tiger Woods - Golfer from USA - Golf ...
Tiger Woods - Golfer from USA - Golf ...
THE GOLF SWING OF TIGER WOODS
American golfer Tiger Woods hits
golfer Tiger Woods viewing a
Woods golfer van het jaar
Retrieved Set

Fig. 1. *Interactive multimedia search paradigm. Query-by-example is used to search database of images. System refines its understanding of search concept from user interaction on this set of results. (Image search results in this figure courtesy of images.google.com.)*

A. Learning Strategies

The content-based learning/search algorithms tasked to solve these problems must operate under three critical constraints.

The first challenge is tackling the *small sample learning* problem [27]. In many practical applications, a user's initial query often gives little information to the system. Algorithms must contend with only a handful of labeled training examples, often coming from unbalanced data sets in high-dimensional feature spaces. Consequently, many potentially discriminating observations go unlabeled.

A second major challenge rises again from the *semantic gap* [10]: correlating user-dependent interpretation of multimedia content with low-level audio-visual descriptors.

Finally, the speed at which both training and search occurs in such systems must be fast. Users do not want to wait indefinitely for search results, and designers must take this into account. These considerations motivate the need for interaction.

B. Interaction Strategy

Though the term implies the learner's subjugation to the user, the true nature of interaction for content-based learners is more self-serving. Interaction will help alleviate some of the challenges inherent to small-sample learning through the extra labeling effort of the user. This labor, however, comes at a cost. The user's time and effort are at a premium, so any extra information the learner wants to gain must be done in the most efficient and useful way. The interaction must be quick and painless for the user while at the same time provide the most information to the learner.

As one can see, the selection of these two components are not mutually exclusive. The learner we choose for multimedia search motivates our choice for interaction strategy and *vice versa*. For the interaction strategy, however, there is a general paradigm that all approaches fit into regardless of the learner: *active learning*. We will motivate this paradigm in the general context of interaction strategies in the next section.

III. GENERAL INTERACTION STRATEGIES

Consider again the retrieval scenario. Using the *query-by-example* approach, a learning algorithm trains itself on a small amount of training data. From this result, the system determines an interaction strategy for the user in order to further improve its idea about the intended target of the search. The system wants to get the best information from the user in order to better learn the user's information need. This is the main goal of the interactive learning strategy.

A. Relevance Feedback

The first interactive learning approach for content-based search that garnered great attention and success was *relevance feedback* [10], [13]–[15]. From the initial returned set of most relevant instances, the user is asked to give explicit feedback by labeling instances in the returned set as being either "relevant" or "irrelevant." This information is then used to refine the search strategy typically by adjusting the notion of similarity between documents in the database. Each round of feedback is intended to bring the system closer to finding the user's implicit target concept.

Initial work in relevance feedback was based on heuristic-based feature reweighting schemes [10], [28] that weighted certain features over others based on the user's past preferences. Since then, a plethora of techniques from different fields have been studied. These include probabilistic frameworks [15], artificial neural

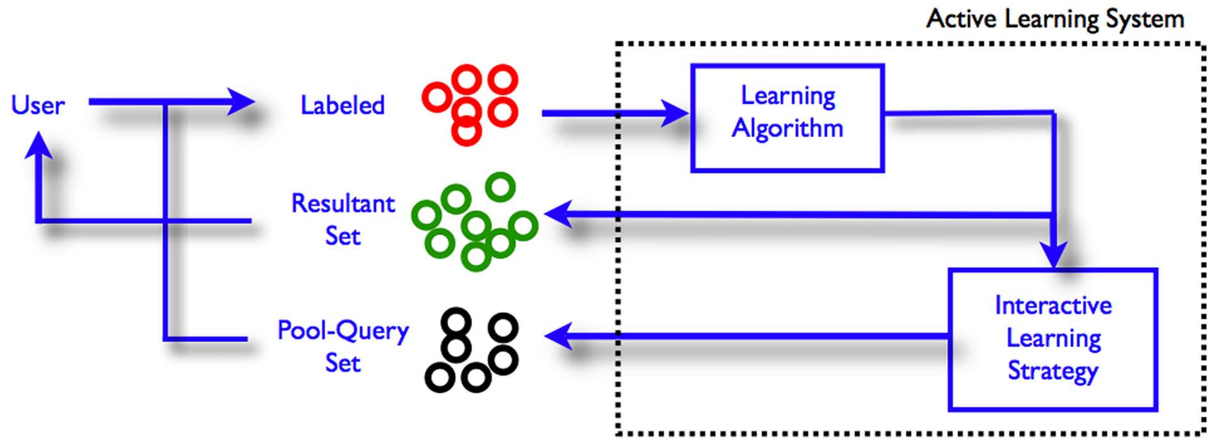


Fig. 2. *Active learning paradigm. Learning and interaction strategies collectively make active learning system. System initially starts with red labeled examples and at each round returns both resultant set (green examples) and to-be-labeled pool-query set (black examples) which when labeled will be added to training set for next round.*

networks [29], and others including those that seek to combine other modalities such as text into the search process [30]. In addition, algorithms from the text retrieval community like query-point-refinement [10] have been adapted to the problem in conjunction with pattern recognition techniques such as discriminant analysis [27], [31].

As we will see in the next subsection, relevance feedback and techniques based on this idea fall under the umbrella of *active learning*, an approach ideally suited to the information retrieval problem.

B. Active Learning

Active learning is a paradigm that proposes ways to incrementally learn from unlabeled data, provided the system has available to it an *oracle*, an entity which knows the correct labeling of all examples [32], [33]. In the case of multimedia information retrieval, this is the user. Given an initial weak retrieval learner, the system asks the oracle to label those points whose correct labeling it deems to be *most informative*: the *pool query set*. The information provided from the pool-query labeling is then used to update the learner and this process can be repeated indefinitely to improve the learning performance of those points in the returned or *resultant set*. This paradigm is illustrated in Fig. 2.

From the perspective of the user, the resultant set is the list of most relevant multimedia documents (with respect to the search concept) and the pool-query set is the set of most informative documents. He must label to "help" the system refine the search. Determining what *most informative* means and how to choose these points for labeling is the fundamental challenge in this field and the main focus of research.

This is active learning in its most general incarnation. Examples of special cases include when the pool-query-set is singular (*stream-based active learning* [34]) or when we are not restricted to unlabeled instances but can seek to label arbitrary points in the feature space (*membership query learning* [35]), among others. In fact, traditional relevance feedback can be seen as a degenerate case of general active learning as the set of top-k returned points serves both as the returned *and* pool query sets.

Active learning is the most natural formalism to the interactive learning problem. Because it is incremental, it is more similar to boosting or sequential training algorithms as opposed to the one-shot training of traditional learning systems. It most closely resembles semi-supervised learning in that its main goal is to properly learn from unlabeled data, though its job is a bit easier in that it has available to it the oracle which can divine the true labeling of unlabeled instances in the data set, as opposed to inferring them as in semi-supervised approaches.[1]

Active learning techniques are used when we encounter two types of constraints in an application area. The first, and historical root of active learning research, arises from expensive data measurement. When new data is expensive or slow to obtain, we want to make sure to choose the most representative, informative training set to model the system. This can be seen in problems as diverse as celestial mechanics [36], statistics [37], and economic theory [38].

The second scenario is when we have a scarcity of labeled data. This is the case where unlabeled examples are plentiful to obtain, but data labeling is time consuming and expensive. This is also referred to as learning with unbalanced data sets and arises in multimedia information retrieval applications.

[1] The distinction between semi-supervised and active learning is blurred a bit with work in selective sampling for query learning in the neural network community.

These two justifications for when to use active learning correlate closely to the historical evolution of the field. Active learning, or active-learning like problems, have been studied in a variety of fields and have a diverse pedigree.

C. Historical Perspective

According to Jaynes [36], the earliest recorded use of active learning in science is attributed to Laplace who used them along with Bayesian reasoning techniques to solve problems in celestial mechanics in the early 19th century. The active learning-like problem of *experimental* or *sequential design* was explored extensively in the fields of statistical and economic theory by Lindley [37] and Federov [38], respectively, in the mid-to-late 20th century.

The field began to mature in the late 20th century as researchers from the pattern recognition and machine learning communities began to explore the potential of these approaches. This was around the same time that researchers began formalizing the organization of active learning approaches.

In the neural network community, several works [32], [35], [39] explored the use of active learning in the context of efficient network training. These techniques fell under a variety of names including *query-learning*, *active learning with membership queries*, and *selective sampling*, though all shared the same underlying approach. In each round of learning, the system chooses a point in feature space and requests its label from the oracle (active learning in which the unlabeled pool of examples consists of all possible points in the low-level feature space). The techniques rely on choosing the two most extreme learner hypotheses given the current set of training data and choosing a point in the feature space where they disagree most for user labeling.

In the context of stream active learning, the query-by-committee system of Freund *et al.* [40] introduced the concept of a *version space* for classification problems. A version space is the set of classifier learners that correctly partition the feature space with respect to the current set of labeled data. Active learning takes place by sampling from this space of consistent hypotheses an even number of classifiers and classifying points as they stream in by committee voting, requesting labels only when the voting results in a tie. The seminal theoretical result of this work is that sampling from the version space in this way decreases its cardinality and that this corresponds to an exponential reduction of the generalization error for classification.

At the same time, general *pool-based* techniques were making headway with systems that provided firm probabilistic analysis on how to choose the optimal unlabeled example(s). In [32], the point(s) with the minimum expected variance of the learner (with respect to the data) was chosen for active labeling in closed form. The work of Roy and McCallum [33] advanced this idea by noting for many learners, the expected variance could not be found in closed-form, and advocated a computationally feasible model to estimate this value using Monte Carlo techniques.

Another seminal theoretical and practical discovery was made by [41] in the context of version space reduction in support vector machine (SVM) classifiers. Exploiting version space duality for SVMs, it was proven that for these classifiers, only unlabeled points that fall within the SVM margin are to be considered for improving classification accuracy. They proved the intuitive result that those unlabeled points closest to the separating hyperplane are the optimal choice to most reduce the size of the version space and thus improve the classification accuracy at each round.

The first two special cases of active learning serve as inspiration for, though are not directly applicable to, the multimedia information retrieval problem. Selective sampling assumes an infinite number of unlabeled points to chose for labeling, which is not the case for a finite, though large, database. Stream-based approaches require users to label unlabeled points at a time, which is impractical from an interface perspective. For these reasons, active learning research for multimedia information retrieval lies in the realm of pool-based approaches.

Applying pool-based active learning techniques to problems of multimedia information retrieval and mining makes sense because the amount of high-quality labeled training data is often dwarfed by the amount of unlabeled data for these application areas, the so-called *small-sample learning* problem. Advances in both textual and visual retrieval problems [41]–[44], face database mining [45], as well as robust video annotation for outlier detection [46] have been made in the recent past using pool-based techniques.

Given these considerations, several interesting questions arise at this point. What form should the learning algorithm take? What is our interactive learning strategy? What is our approach or criterion for best utilizing the user's effort in interactive labeling? We will seek to address these questions and profile active learning research in multimedia retrieval in the following sections. We first begin by outlining classes of learning algorithms used for active learning in multimedia information retrieval applications.

IV. LEARNING ALGORITHM

In general, the learning algorithm chosen for a particular problem helps to dictate the exact implementation of the interactive learning strategy. More clearly, each unique learner will change our definition of *most informative* point(s) and how we go about choosing it.

Because of our application considerations (small-sample learning, semantic gap, and speed) we have to restrict our discussion of learning algorithms to those that are robust to these constraints. We next profile three important classes of learners from the spectrum of techniques used in active learning systems.

A. Classification-Based Algorithms: Support Vector Machines

Often multimedia information retrieval applications can be formulated as classification problems. Whether we are modeling a scenario as a binary classification problem (relevant versus irrelevant images [41] or anomalous versus normal video events [47]) or whether we are looking at multiclass discrimination problems such as the organization and search of news group messages [32], we can rely on classifier-based active learning techniques to help solve our problems. Though many classification algorithms have been explored for active learning, we will focus on an algorithm that has had great applicability to multimedia information retrieval applications in recent years: SVMs.

SVMs: Consider a binary classification problem given a set of labeled training data $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$ where samples $\mathbf{x}_i \in \mathbb{R}^d$ and binary labels are given as $y \in \{-1, 1\}$. In the case of multimedia information retrieval, we can consider $\mathbb{R}^d$ the d-dimensional space of low-level features so that each image or video has a unique feature-vector descriptor. The labels y_i are markers indicating whether examples are either relevant or irrelevant. Our goal is to learn a linear function that will separate relevant examples from irrelevant ones. In doing so, we can separate the documents that are relevant to the user's search from ones that are not.

Consider also that there is a mapping between the original input space $\mathcal{X}$ to a higher (possibly infinite) dimensional feature space $\mathcal{F}$ given by $\Phi : \mathcal{X} \to \mathcal{F}$. (This is appealing because many problems which are not linearly separable in $\mathcal{X}$ become linearly separable when mapped to $\mathcal{F}$.)

In their most general form, SVMs are classifiers of the type

$$f(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i \mathcal{K}(\mathbf{x}_i, \mathbf{x}) \tag{1}$$

where $\hat{y} = 1$ if $f(\mathbf{x}) \geq 0$ and $\hat{y} = -1$ if $f(\mathbf{x}) < 0$, $\{\alpha_i\}_{i=1}^N \in \mathbb{R}^N$, and $\mathcal{K}(\mathbf{x}_i, \mathbf{y}_j)$ is a *Mercer kernel* satisfying the property $\mathcal{K}(\mathbf{x}_i, \mathbf{x}_j) = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle$. The use of such kernels allows for the implicit projection of the input space $\mathcal{X}$ to a higher space $\mathcal{F}$. This is better seen if we rewrite (1) as

$$f(\mathbf{x}) = \mathbf{w}^T \Phi(\mathbf{x}) \tag{2}$$

where $\mathbf{w} = \sum_{i=1}^{N} \alpha_i \Phi(\mathbf{x}_i)$ and $\mathbf{w} \in \mathcal{F}$. In this form, it is easy to see that SVMs most simply are hyperplanes in the

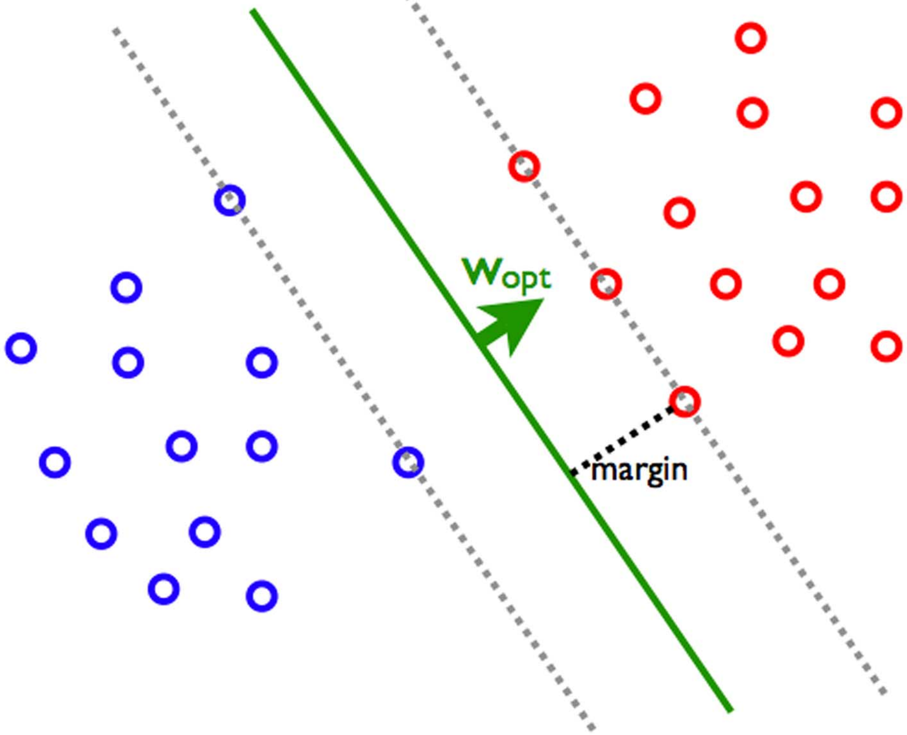


Fig. 3. ***SVM classification problem. Goal is to find classifier with largest margin between closest positive and negatively labeled examples: support vectors. Normal vector for optimal separating hyperplane $\mathbf{w}_{\text{opt}}$ is found using a quadratic optimization procedure.***

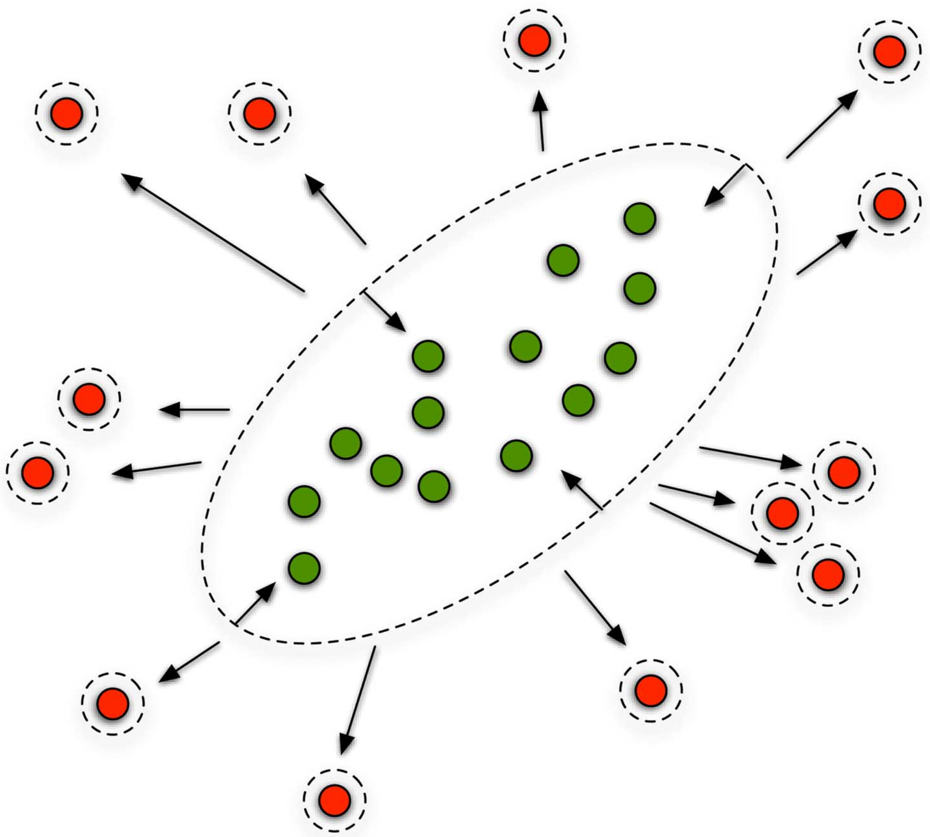


Fig. 4. *BDA problem. Green circles represent positive examples, red circles represent negative examples. BDA seeks to transform feature space so that positive examples cluster close together and each negative instance is pushed away as far as possible from positive cluster.*

feature space $\mathcal{F}$. The $\{\alpha_i\}_{i=1}^{N}$ needed to compute the normal vector $\mathbf{w}$ of this hyperplane are obtained by maximizing the minimum distance of all labeled points from the hyperplane. The final value of this distance is referred to as the *margin* and those labeled points that lie on the margin are referred to as the *support vectors*. This is illustrated in Fig. 3. The margin can be normalized so that the absolute value of $f(\mathbf{x})$ for support vectors is always one. This optimal hyperplane is found via a quadratic optimization procedure. For a more detailed treatment, the reader is referred to [48].

B. Query-Point Refinement Algorithms: Biased Discriminant Analysis

Query-point refinement is a pseudo-ranking learning strategy that is closely coupled with interactive learning. That is, its greatest usefulness comes through interaction with a user. Historically, these techniques have been closely associated with the special case of active learning: relevance feedback. On its own, it is closest to a weighted k-nearest neighbor ranker. Though many other query-point-refinement learners and techniques exist [49]–[51], for the purposes of this paper, we focus on one ideally suited for the small-sample learning problem: Biased Discriminant Analysis (BDA) [27].

BDA was developed to address the inherent problem in traditional feature reweighting techniques that try to cluster all negatively (irrelevant) labeled examples together. These approaches do not make intuitive sense because negative examples can come from many different classes and many different parts of the feature space. (Irrelevant Google Image Search returns may, as a group, have nothing to do with one another.)

Accordingly, BDA casts the problem of relevance feedback from a two-class (positive and negative) to a one-to-many class (one positive, multiple negative) problem. The idea is that positive examples are derived from one class while negative examples may come from multiple classes.

The goal, as illustrated in Fig. 4, is to find a feature space transformation that closely clusters the positive examples while pushing away the negative ones. In its full form this becomes a query-point-refinement algorithm. Each round of user feedback results in a new set of labeled points, which in turn yields a new BDA transformation of the feature space, which results in the centroids of both the positive and negative examples being moved.[2]

The BDA problem is characterized by the following objective function:

$$\mathbf{W}^* = \arg\max_{\mathbf{W}} \left| \frac{\mathbf{W}^T \mathbf{S}_{PN} \mathbf{W}}{\mathbf{W}^T \mathbf{S}_P \mathbf{W}} \right| \tag{3}$$

where $\mathbf{S}_P$ is the intraclass scatter matrix for positive examples and $\mathbf{S}_{PN}$ is the interclass scatter matrix between positive and negative examples. Specifically

$$\mathbf{S}_P = \sum_{\mathbf{x} \in \mathrm{P}} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{P}})(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{P}})^T \tag{4}$$

$$\mathbf{S}_{PN} = \sum_{\mathbf{y} \in \mathrm{N}} (\mathbf{y} - \boldsymbol{\mu}_{\mathbf{P}})(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{P}})^T \tag{5}$$

where $\mathbf{x}$ and $\mathbf{y}$ are feature points from the positive and negative labeled sets P and N, respectively, and $\boldsymbol{\mu}_{\mathbf{P}}$ is the mean of the points in P.

As in traditional discriminant analysis techniques such as Linear Discriminant Analysis (LDA), and its general form Multiple Discriminant Analysis (MDA), the solution reduces to solving the generalized eigenvalue problem for the Rayleigh quotient in (3). The columns of matrix $\mathbf{W}$ correspond to the generalized eigenvectors corresponding to the largest eigenvalues

$$\mathbf{S}_{PN} \mathbf{w}_i = \lambda_i \mathbf{S}_W \mathbf{w}_i. \tag{6}$$

Once these are found, the discriminating transformation matrix is computed as

$$\mathbf{A} = \boldsymbol{\Phi} \Lambda^{1/2} \tag{7}$$

where $\boldsymbol{\Lambda}$ is the diagonal eigenvalue matrix and $\boldsymbol{\Phi}$ is the corresponding eigenvector matrix so that $\mathbf{W}^* = A^T A$. The distance between two points then becomes

$$\mathrm{d}(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{W}^* (\mathbf{x}_i - \mathbf{x}_j). \tag{8}$$

This distance corresponds to the canonical euclidean distance in the new feature space induced by the transformation $\mathbf{A}$.

All unlabeled points in the database can be rank sorted with respect to this distance measure and returned to the user as the result of the retrieval process.

In the next section, we look to the use of a ranker-learning framework to the information retrieval problem.

C. Ranker-Based Algorithms: Bipartite Ranking

Though there has been little interest in this direction until recently, problems of information retrieval, and particularly relevance feedback-style scenarios, can be posed in a ranking framework. Consider the case where the user has a number of multimedia documents that she has labeled. This labeling induces a ranking on the set of labeled instances: positively labeled instances being preferred over the negatively labeled ones. A ranking function can then be constructed for all the images in the database so that the images we have labeled as positive are ranked above those that have been labeled negative, the hope being that unlabeled images similar to the labeled ones will be ranked in the same way. An initial work in this regard was [52]. These ideas can be illustrated in Fig. 5.

Despite the wealth of research into ranking problems, *active ranking* is a new area. That is, to the best of our knowledge, there has been only one work [53] studying how to choose the most informative unlabeled points with respect to the ranking scenario.

In this section, we will outline this approach to looking at active learning in the context of rankers along with a review of the bipartite ranking scenario.

Bipartite Ranking: Consider the input to a learning algorithm is a set of training examples $S = \{(x_i, y_i)\}_{i=1}^N \in (\mathcal{X} \times \mathcal{Y})^N$. The learning algorithm should predict if the new instance is relevant or irrelevant by learning from S a function $h : \mathcal{X} \rightarrow \mathcal{Y}$ so that given a new instance $x \in \mathcal{X}$, the algorithm would predict the label as $h(x)$. In the case where $\mathcal{Y}$ is $\{0,1\}$, the learning algorithm is a standard binary classification algorithm as before. However, in our case, we do not need a labeling of the instances as relevant or irrelevant as in traditional multimedia information retrieval techniques. Instead, we desire an ordering of the instances in such a way that relevant instances are on the top of the list and the irrelevant ones are at the bottom. Such an ordering problem corresponds to the bipartite ranking formulation introduced in [54] and analyzed further in [55] and [56].

The goal of ranking is to obtain an ordered list of entities where order is determined by preference or choice. The preference is either hand-crafted or more desirably learned from the annotated data. Learning a ranker amounts to finding an axis in the feature space that data points are mapped to so that the relative position between points reflects the desired preference. The absolute value of the projected examples does not have any particular meaning and this feature distinguishes ranking from ordinal regression. Further, absolute rankings do not

[2] The BDA transformation matrix can also be used as a feature space reduction technique akin to Principal Component Analysis (PCA).

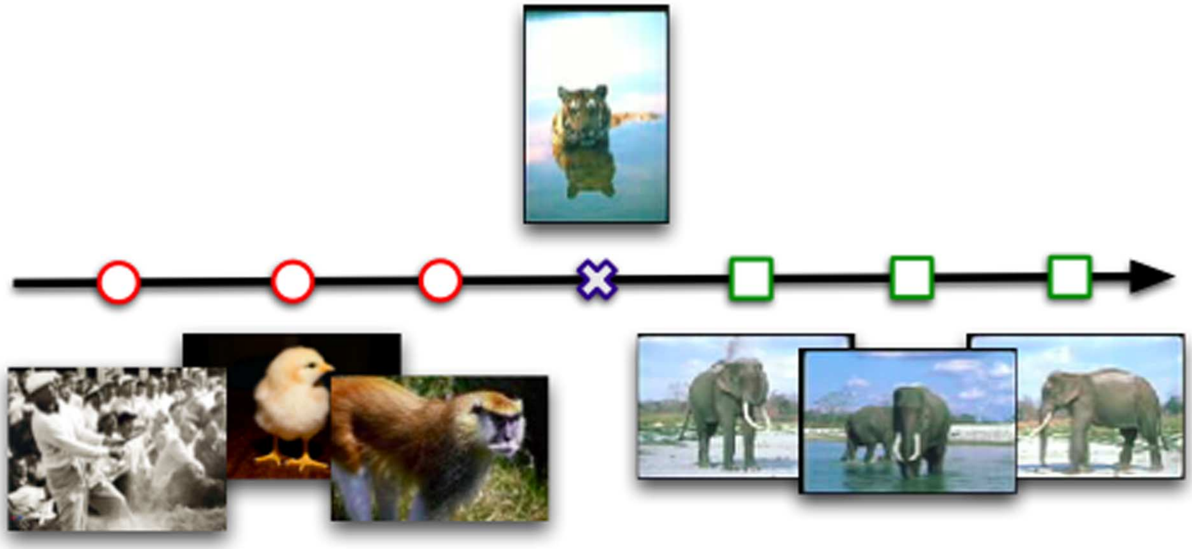


Fig. 5. *Queried concept is an elephant in a blue background. Plot shows ranker output on a real line in which squares indicate relevant instances and circles indicate irrelevant instances. Cross mark refers to unlabeled instance which falls between current labeled set of relevant and irrelevant instances leading to confusion about its label, an idea that is exploited later to motivate the interactive strategy for active ranking.*

need to have any particular meaning across disjunctive data sets. Ranking is solely based on the relative position of the one dimensional mapping.

Consider again inputing to the learning algorithm a set of training examples $S = \{(x_i, y_i)\}_{i=1}^{N} \in (\mathcal{X} \times \mathcal{Y})^N$ where $\mathcal{X} = \mathbb{R}^d$ and $\mathcal{Y} = \{0, 1\}$. Formally, we require that the algorithm learn from S a real-valued ranking function $f : \mathcal{X} \rightarrow \mathbb{R}$ that assigns scores to instances and thereby induces an ordering over $\mathcal{X}$: an instance $x \in \mathcal{X}$ is ranked higher by f than an instance $x' \in \mathcal{X}$ if $f(x) > f(x')$ and lower if $f(x) < f(x')$.

In the bipartite ranking setting [54], the quality of a ranking function is measured as

$$\hat{R}(f; S) = \frac{1}{n_0 n_1} \sum_{\{i:y_i=0\}} \sum_{\{j:y_j=1\}} \mathbf{I}_{\{f(x_j) \le f(x_i)\}} \tag{9}$$

where $n_l = |\{i : y_i = l\}|$ and $\mathbf{I}_{\{\cdot\}}$ denotes the indicator variable whose value is one if its argument is true and zero otherwise. The bipartite ranking error effectively counts an error each time a relevant instance (label 1) is ranked lower by f than an irrelevant instance (label 0), represented by $\mathbf{I}_{\{f(x_j) \le f(x_i)\}}$.

In [54], Freund *et al.* introduced a boosting style algorithm called RankBoost for solving the ranking problem. It is similar to traditional boosting, except the boosting is done over rankers, as opposed to classifiers. For the sake of brevity, we omit this analysis, but we direct interested readers to [52]–[54] for more information.

V. INTERACTION STRATEGY

Often, in order to properly choose unlabeled points to form the pool-query set, we must take into consideration the learning strategy we have adopted. That is, we need to understand how best to form this set under the constraint of the learner being considered. In this section, we outline interactive learning strategies for the selection of points for inclusion in the pool-query set.

A. Classifier-Based Strategies: SVM Version Space Reduction

In the case of support vector machines, finding the optimal set of unlabeled points for the user to label is motivated by version space reduction. We begin this analysis by first motivating version spaces.

Version Space: The general problem of classifier selection boils down choosing the "best" classifier among all classifiers that correctly classify the training data. This set of classifiers is known as the *Version Space* [57].

A subtle point here is that we are implicitly making the overoptimistic assumption that the "true" classifier for a particular multimedia search concept exists in the set of consistent classifiers. This is a necessary assumption, however, because each unique user and search session can yield a different target concept and as such will result in different "true" classifiers that may not always reflect the underlying distributions inherent in the data.

If we restrict the discussion to linear SVM classifiers and adapting the naming convention in Section IV-A, there exists a one-to-one correspondence between classifiers and their normal vectors $\mathbf{w}$. The version space can therefore be written as

$$\mathcal{V} = \left\{ \mathbf{w} \in \mathcal{F} : \|\mathbf{w}\| = 1, \left\{ y_i\left(\mathbf{w}^T \Phi(\mathbf{x}_i)\right) \geq 0 \right\}_{i=1}^{N} \right\}. \quad (10)$$

In other words, for optimization we restrict our discussion to those normal vectors with unit norm that correctly classify the training data. The version space is more than a statement of the constraint set for SVM optimization. It is the foundation upon which we can build a dual view of the classification problem.

Consider a potential parameter space $\mathcal{W}$ where $\mathbf{w} \in \mathcal{W}$ and $\|\mathbf{w}\| = 1$. Because of the unit norm constraint, this corresponds to points on the surface of the unit hypersphere in $\mathcal{W}$. Consider now the (slightly rewritten) set of labeled data constraints $\{(y_i\Phi(\mathbf{x}_i))^T\mathbf{w} \geq 0\}_{i=1}^{N}$. Since we are in the parameter space $\mathcal{W}$, each vector $y_i\Phi(\mathbf{x}_i)$ can now be seen as a normal vector of a hyperplane in $\mathcal{W}$ (assuming $\|\Phi(\mathbf{x})\| = 1$, which for many kernels is the case). This is a critical step, because now each constraint $(y_i\Phi(\mathbf{x}_i))^T\mathbf{w} \geq 0$ corresponds to a half-space in $\mathcal{W}$. The intersection of all these half-spaces and the unit hypersphere results in a delineation of the hypersphere surface which corresponds exactly to the set of points given in (10).

In this dual space, SVM optimization corresponds to growing the largest hypersphere which can be inscribed within the delineated space in $\mathcal{W}$, whose center lies on the version space, i.e., the remaining surface of the unit hypersphere. The radius of this inscribed hypersphere corresponds to the SVM margin, and the labeled sample hyperplanes it touches at the boundaries of the version space are the support vectors. (A more detailed treatment can be found in [41].)

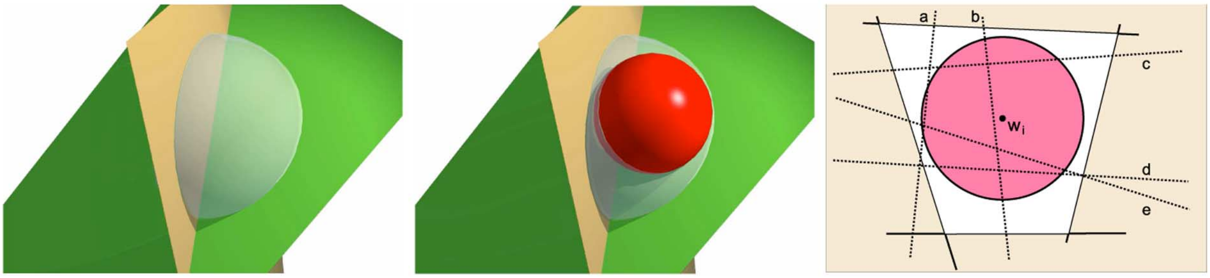
The version space is critical for active learning with SVMs. Each unlabeled sample point corresponds to a hyperplane which passes through the unit hypersphere in the parameter space. Some unlabeled hyperspheres may also pass through the current version space. A labeling of these points would decrease the size of the version space, fine tuning the classification and thus improving performance. This can be seen in Fig. 6.

Version Space Reduction: Each unlabeled training example has the potential to reduce version space size once it has been labeled, thus reducing the number of possible classifiers that will properly classify the data. Doing this is not just intuitively attractive. It has been proven by Sueng *et al.*'s query by committee (QBC) work that halving the size of the version space each round exponentially reduces the classifier's generalization error (relative to random sampling) [40].

Despite this result, it is still computationally infeasible to calculate the reduction in version space size following the two possible labelings ({−1,1}) of all unlabeled points which exist in the version space. Instead, an intelligent approximation is needed. In their seminal paper [41], Tong and Koller observed the weight vector $\mathbf{w}$ found by SVM optimization approximates the center of mass of the version space, the so-called *Bayes point*. They reasoned if at each round if the active learner chooses those unlabeled points whose hyperspheres pass closest to the classification hyperplane, the version space size would be approximately halved at each round of active learning.

B. Query-Point Refinement-Based Strategies: Diversity Analysis

Active learning algorithms relying on query-point refinement most closely relate to traditional relevance feedback. As more unlabeled data is labeled in the relevance



(a)

(b)

(c)

Fig. 6. ***(a) Version space duality. Surface of hypersphere represents unit weight vectors. Each of the two hyperplanes corresponds to a labeled training instance. Each hyperplane restricts area on hypersphere in which consistent hypotheses can lie. Here, version space is surface segment of hypersphere closest to the camera. (b) SVM classifier in version space. Dark embedded sphere is largest radius sphere whose center lies in version space and whose surface does not intersect with the hyperplanes. Center of embedded sphere corresponds to SVM, its radius is the margin of the SVM in $\mathcal{F}$, and training points corresponding to the hyperplanes that it touches are the support vectors. (c) Top down view, dotted lines correspond to unlabeled point hyperplanes. Choosing those points that roughly bisect the version space (are closest of $\mathbf{w}_{\rm opt}$) help reduce version space size and improve performance.***

feedback process, the optimal transformation of the feature space continues to change, thus allowing for the refinement of the initial query point with respect to the original feature space. We can lift query-point refinement out of the degenerate case of active learning (the relevance feedback paradigm) and choose for the user to label a distinct pool-query set as in [44] and [58]. In particular, we advocate the incorporation of diversity information into the labeling process.

As motivation, consider that whether a traditional or active learning-based paradigm is used, the user is often asked to label examples which are quite similar to one another, often times as a result of examples clustering in the same area of the feature space. In a small-sample setting, especially when we want to minimize both user and system effort, it makes more sense for the user to label a *diverse* set of points for each pool query rather than many similar points which are, by comparison, much less informative.

From a purely active learning viewpoint, one of the first works to incorporate diversity sampling was [43] and subsequently [42], where the notion of angular diversity was investigated for SVMs and as such using diversity is not unique to query-point refinement techniques. The idea of using angular diversity was motivated specifically by the version-space reduction requirements inherent in SVM active learning. Additionally, the use of information-theoretic diversity sampling has recently been used for a variety of active learners, including ranker-based [53], SVM-based [59], and query-point refinement-based [58] techniques.

Incorporating Diversity: A general active learning algorithm chooses both a resultant and pool-query set to present to the user at each step. We assume that the algorithm narrows down the set of all unlabeled points at each round of feedback to a *candidate pool-query set* C. In a query-point refinement algorithm, this set can be localized to a neighborhood of unlabeled examples around the query-centroid by either heuristic or index-based nearest neighborhood techniques. Once a candidate pool-query set has been found, the goal then becomes selecting a *diverse*, or representative, set of points from this larger set C, to include in our pool-query set, P.

Before we elaborate on diversity measures used to accomplish this task, we first address a practical issue. Assume at each round the cardinality of the candidate-pool set remains the same and is L. Assume also that we wish the cardinality of the pool-query set to be K where $K \ll L$. There then becomes $\binom{L}{K}$ possible representative sets from which the system must choose the most representative, or diverse, set of unlabeled instances for the user to label. Even for moderate sample sizes, however, this number becomes quickly intractable. Clearly, greedy algorithms must be used along with our diversity measure. In recent work, two main diversity measures have been used: angular and entropic.

Angular Diversity: Angular diversity was first introduced in the context of active learning for SVMs in [43] and later used in a query-point-refinement setting in [44]. Given the set of unlabeled instances in the candidate set C, $\{\mathbf{x}_i\}_{i=1}^{L}$ ($\mathbf{x}_i \in \mathbb{R}^d$), the angular diversity between any two instances $\mathbf{x}_i$ and $\mathbf{x}_j$ can be defined as

$$\cos\left(\angle(\mathbf{x}_i, \mathbf{x}_j)\right) = \left| \frac{(\mathbf{x}_i - \mathbf{x}_c)^T (\mathbf{x}_j - \mathbf{x}_c)}{\|\mathbf{x}_i - \mathbf{x}_c\| \|\mathbf{x}_j - \mathbf{x}_c\|} \right| \tag{11}$$

where $\mathbf{x}_c$ is the mean of the relevant instances. A diverse set can then be incrementally constructed in a greedy fashion by minimizing

$$\max_{l \in P_n} \cos\left(\angle(\mathbf{x}_l, \mathbf{x}_j)\right) \tag{12}$$

where P_n is the current pool-query set (greedy-increment round n) and $\mathbf{x}_j$ is an instance from the candidate set C under consideration for addition to the updated pool-query set P_{n+1}. In addition to being angularly diverse, we also require instances to be sufficiently close to the query centroid $\mathbf{x}_c$. Accordingly, the final cost function for each instance in the candidate set becomes

$$F(\mathbf{x}_i) = \alpha d(\mathbf{x}_i, \mathbf{x}_c) + (1 - \alpha) \max_{l \in P_n} (\cos \angle(\mathbf{x}_l, \mathbf{x}_i)) \tag{13}$$

where $\mathbf{x}_i \in C$ and α denotes a convex mixing parameter between diversity and centroid proximity. The instance among all C with the smallest value for the final cost function is chosen and added to the new increment of the pool-query set P_{n+1}. This process is repeated until all greedy increment rounds are completed and the final pool-query set is complete.

Information-Theoretic Diversity Sampling: A drawback of using the convex cost function as above is knowing how to set the tradeoff parameter α. To combat this, [53] took a slightly different approach to forming a cost function and that is through the use of information-theoretic diversity. Associating high entropy with diversity is intuitively attractive as entropy is essentially a measure of randomness in a variable.

As pointed out in [60], two samples are enough to estimate the entropy of a density. The first sample is used to estimate the density and the second sample is used to estimate the entropy. That is, the system must choose K instances which are diverse and representative of the L instances in candidate set C. In other words, the problem reduces to identifying K points which are used to estimate the density in such a way that the entropy estimated over the remaining $L - K$ points is maximized. In practice, density estimation is typically done using

Parzen windowing techniques and empirical entropies through numerical integration approaches.

We note again that picking the optimal pool-query set is computationally infeasible and hence we must resort to a greedy algorithm that begins by selecting the unlabeled instance that is closest to the query centroid. Subsequently, instances are incrementally added to the pool query set such that their addition maximizes the entropy computed with respect to the instances in $C \setminus P_n$, where P_n again represents the current pool query set at round n.

C. Ranker-Based Strategies: Ranking Clarity

As we have seen from previous sections, in the case of SVM active learning, the candidates for the pool-query set correspond to the unlabeled points which lie in the version space. In a query refinement algorithm, one can choose from a large number of points in the neighborhood of the query centroid. In this section, we look at how best to choose a collection of points for labeling in the bipartite ranking scenario.

In general, the pool-query set is chosen as those instances that are hardest to handle or most confusing for the current classifier/ranker. We rely on a quantity called the *clarity index* for each unlabeled instance in order to represent this idea.

Let $T = ((x_1, y_1), \ldots, (x_N, y_N))$ be the complete set of labeled instances obtained from previous active learning rounds and f be the current ranker. For every unlabeled instance x_i^u *relevance loss* $RL(x_i^u, f, T)$ is defined as

$$RL\left(x_i^u, f, T\right) := \frac{1}{n_0}\left|\left\{j : f\left(x_i^u\right) \leq f(x_j), y_j = 0\right\}\right| \qquad (14)$$

and *irrelevance loss* $IL(x_i^u, f, T)$ as

$$IL\left(x_i^u, f, T\right) := \frac{1}{n_1}\left|\left\{j : f\left(x_i^u\right) > f(x_j), y_j = 1\right\}\right|. \qquad (15)$$

Relevance loss can be interpreted as the bipartite ranking loss $\hat{R}(f; S_R)$ [defined in (9)] where the set S_R is given by $((x_i^u, 1), T_I)$ where T_I represents the irrelevant instances present in the set T. Irrelevance loss is given by the bipartite ranking loss $\hat{R}(f; S_I)$ [defined in (9)] where the set S_I is given by $((x_i^u, 0), T_R)$ where T_R represents the relevant instances present in the set T.

By definition of the bipartite ranking loss, a good ranking function is expected to have low relevance loss for relevant instances and low irrelevance loss for irrelevant instances. The *clarity index* of an unlabeled instance x_i^u with respect to a ranking function f and labeled set T is

$$CI\left(x_i^u, f, T\right) := \left|RL\left(x_i^u, f, T\right) - IL\left(x_i^u, f, T\right)\right|. \qquad (16)$$

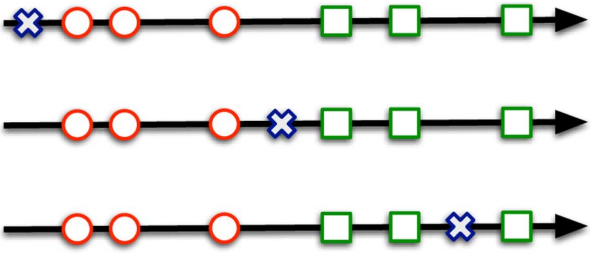


Fig. 7. *Green squares and red circles represent ranking function evaluated at relevant and irrelevant instances, respectively. Blue cross indicates ranking function evaluated on an unlabeled instance. It is clear that the top-most case is the easiest, followed by the bottom-most case, and the middle one is hardest. Relevance loss = 1; Irrelevance loss = 0; Clarity index = 1 (top). Relevance loss = 0; Irrelevance loss = 0; Clarity index = 0 (center). Relevance loss = 0; Irrelevance loss = 2/3; Clarity index = 2/3 (bottom). Difficulty in ranking is captured by clarity index values.*

Clearly, the clarity index orders the instances in terms of their difficulty for the ranking function. The higher the clarity index, the easier it is to classify an instance. A simple illustration is presented in Fig. 7.

The clarity index is evaluated for every unlabeled instance and the instances with the L smallest clarity index values form the candidate set C for the pool query set.

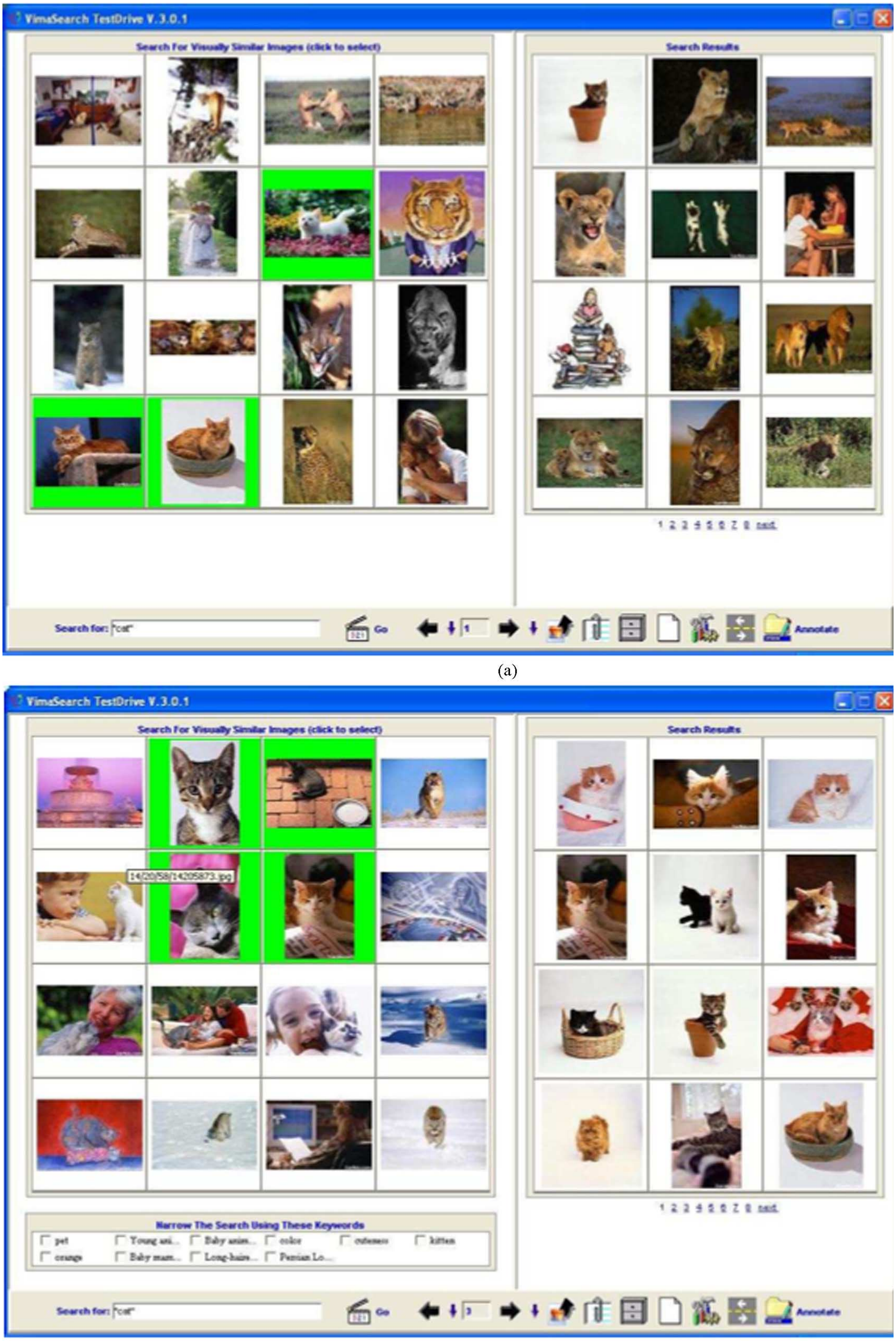
VI. APPLICATIONS

In this section, we profile a selection of results and applications based on techniques motivated in previous sections. These are intended to give the reader a sampling of results balanced between application domain and the active learning approach and are by no means an authoritative review. For brevity, we have omitted a specific profiling of results for query-point refinement instead choosing to profile them as a component technology in the ranking-based approach in Section VI-C. The interested reader is referred to [44] and [58] for an in-depth treatment. We begin by profiling results from the classification algorithm perspective.

A. SVM Active Learning for Image Search

A straightforward application of active learning techniques in multimedia retrieval is the image search domain. The SVM active learning approach can be adapted in a straightforward way, mapping low-level visual features to high-level semantic search concepts.

To illustrate the learning experience from the user's perspective, we present a sample query session to demonstrate how a query concept is learned in the active learning framework in Fig. 8. The user interface shows two frames. The frame on the left-hand side is the feedback frame, on which the user marks images in the pool query set as relevant or irrelevant. On the right-hand side,



(a)

(b)

Fig. 8. *Successive rounds of active learning for query "cat." As can be seen, as images are labeled for the pool-query set from feedback panel, accuracy of the returns in resultant set, return panel, increases. Screenshots are for Rounds two and six respectively.*

the search engine returns what it considers matching the concept learned this far from the image database, the resultant set.

Given an initial labeling of cat images in the feedback frame, the SVM active learning algorithm refines the classification boundary between "cat" and "non-cat" images and then returns the second screen in Fig. 8(b). In this figure, we can see that the results in the (right-hand side) result frame have been greatly improved.

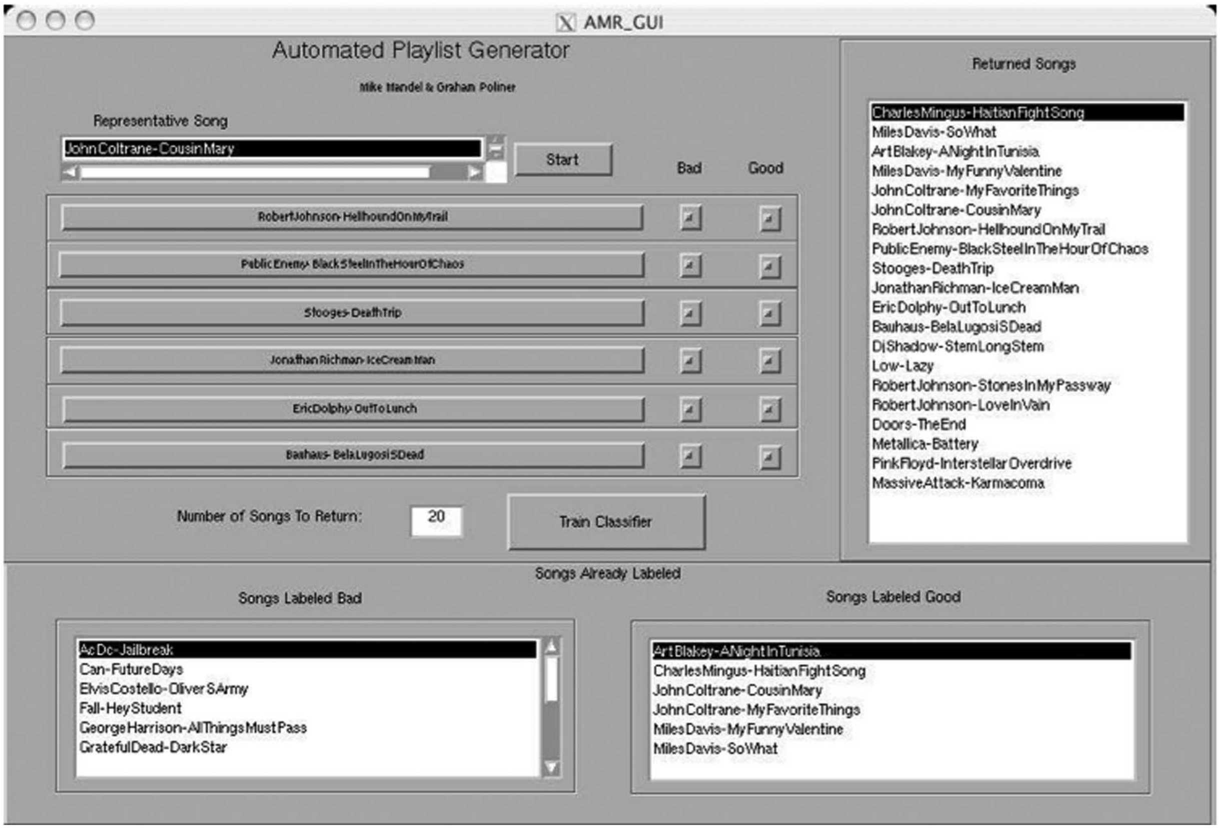
The performance of this system was evaluated using a large collection of images using multiresolution low-level features such as color and texture. The databases investigated were four, ten, and 15 category image sets from the COREL image database [61]. Experimental results comparing the SVM active learning approach showed its viability in comparison with traditional query-point refinement techniques for relevance feedback as well as passive SVM learning. In addition, it was found that incorporating diversity in SVM active learning resulted in greater improvement for conceptually complex datasets. For more details, the reader is referred to [41] and [42].

In later adaptations of this paper [42], [62], angular diversity information as well as *perceptual concept detection*, disambiguating keywords through active learning strategies using images, was implemented to improve the performance of these learners with respect to the initial results. This work was also adapted into a commercial offering known as ImageBeagle [63].

B. SVM Active Learning for Music Search

Music information retrieval is another domain well suited to the use of active learning techniques. Specifically, music search and playlist generation may be cast as active learning problems. In both of these cases, active learning maps low-level audio features such as timbre or spectral shape [64] to higher-level concepts like genre, mood, and style.

An example of an active learning interface for music search can be seen in Fig. 9. The pool-query set takes up the upper left section of the interface along with check boxes for rating whether or not each song is appropriate to the query. Any of the songs can be played by clicking on its name. The right side of the interface shows the resultant set, and the bottom two panes show the labeled songs. In Fig. 9, the user is in the process of searching for the genre "jazz."



Representative Song: JohnColtrane-CousinMary

RobertJohnson-HellhoundOnMyTrail
PublicEnemy-BlackSteelInTheHourOfChaos
Stooges-DeathTrip
JonathanRichman-IceCreamMan
EricDolphy-OutToLunch
Bauhaus-BelaLugosiSDead

Number of Songs To Return: 20

Returned Songs:

CharlesMingus-HaitianFightSong
MilesDavis-SoWhat
ArtBlakey-ANightInTunisia
MilesDavis-MyFunnyValentine
JohnColtrane-MyFavoriteThings
JohnColtrane-CousinMary
RobertJohnson-HellhoundOnMyTrail
PublicEnemy-BlackSteelInTheHourOfChaos
Stooges-DeathTrip
JonathanRichman-IceCreamMan
EricDolphy-OutToLunch
Bauhaus-BelaLugosiSDead
DjShadow-StemLongStem
Low-Lazy
RobertJohnson-StonesInMyPassway
RobertJohnson-LoveInVain
Doors-TheEnd
Metallica-Battery
PinkFloyd-InterstellarOverdrive
MassiveAttack-Karmacoma

Songs Labeled Bad:

AcDc-Jailbreak
Can-FutureDays
ElvisCostello-OliverSArmy
Fall-HeyStudent
GeorgeHarrison-AllThingsMustPass
GratefulDead-DarkStar

Songs Labeled Good:

ArtBlakey-ANightInTunisia
CharlesMingus-HaitianFightSong
JohnColtrane-CousinMary
JohnColtrane-MyFavoriteThings
MilesDavis-MyFunnyValentine
MilesDavis-SoWhat

Fig. 9. ***Graphical user interface for training phase of music playlist generator. User begins by entering a song in query-field on the upper-left corner of window. Results are returned in vertical panel on right, and to-be-labeled pool-query instances are listed below search query-field. Running list of previously labeled songs is maintained in panel frames on the bottom of the window.***

An example of an active learning interface for playlist generation can be seen in Fig. 10. Specifically, relevance feedback is particularly well suited for playlist generation in a music player because a single pool of results is generated. The user's normal interaction with the player provides training labels and feedback. For example, if a user listens to a song in its entirety, it can be given a positive label. If an inappropriate song comes on, the user can skip it, assigning it a negative label. Fig. 10 illustrates a user building a playlist for the genre "rap."

The performance of the SVM active learning approach was compared to traditional SVM learning on the classification of 1210 pop songs from the USPOP dataset [65]. Songs were classified by genre, mood, and style as determined by the allmusic online music guide [66]. Each genre, mood, or style was used separately to train and test classifiers and the results were measured in terms of classification accuracy for all remaining songs and the precision of the top 20 songs. Active learning provided significantly improved results, halving the number of training examples needed to achieve a given level of accuracy. For the same number of labeled examples, active learning resulted in a 10% increase in precision-at-20. For more information, the reader is referred to [24].

C. Active Ranking for Image Search

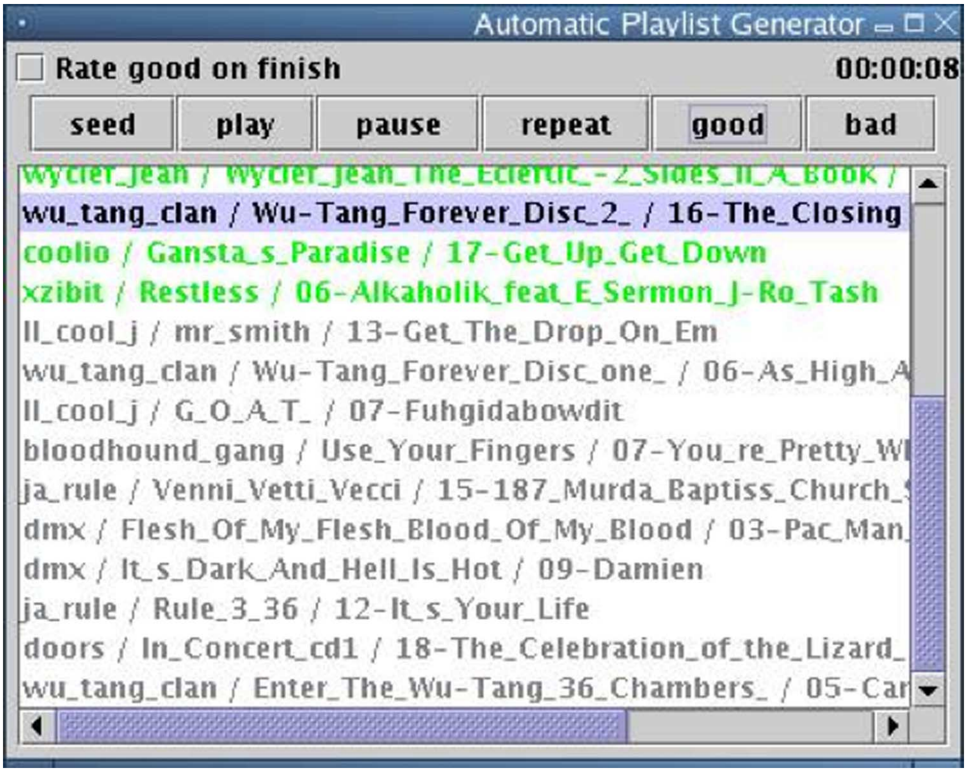
An experimental setup similar to the application of SVM active learning to the image search problem was explored for active ranking [53]. To explore the practical performance of the diverse active ranking system, extensive experimentation was performed using a 5000 image subset of the COREL image database. To appropriately model the small sample learning scenario, only 1400 images were used for target sets.

The first, second, and third moments in each channel of the HSV color space, first and second wavelet sub-band moments at three levels of decomposition, and a Water-filling algorithm were used for color, texture, and shape features, respectively. In total, a 47-dimensional feature vector was extracted from each image.

The system derived its initial ranking function via RankBoost with BDA as the weak ranker. It returned to the user both the K similar images and the pool-query set of images to label for active ranking in each of the following incarnations:

Random Active Ranking Bipartite Ranking: with one set of labeled data, and no further user labeling;

Plain Active Ranking: active ranking without using diversity information (asking the user to label those instances with lowest Clarity Index);



wyclef_jean / Wyclef_Jean_The_Eclectic_-_2_Sides_II_A_Book /
wu_tang_clan / Wu-Tang_Forever_Disc_2_ / 16-The_Closing
coolio / Gansta_s_Paradise / 17-Get_Up_Get_Down
xzibit / Restless / 06-Alkaholik_feat_E_Sermon_J-Ro_Tash
ll_cool_j / mr_smith / 13-Get_The_Drop_On_Em
wu_tang_clan / Wu-Tang_Forever_Disc_one_ / 06-As_High_A
ll_cool_j / G_O_A_T_ / 07-Fuhgidabowdit
bloodhound_gang / Use_Your_Fingers / 07-You_re_Pretty_Wl
ja_rule / Venni_Vetti_Vecci / 15-187_Murda_Baptiss_Church_
dmx / Flesh_Of_My_Flesh_Blood_Of_My_Blood / 03-Pac_Man_
dmx / It_s_Dark_And_Hell_Is_Hot / 09-Damien
ja_rule / Rule_3_36 / 12-It_s_Your_Life
doors / In_Concert_cd1 / 18-The_Celebration_of_the_Lizard_
wu_tang_clan / Enter_The_Wu-Tang_36_Chambers_ / 05-Car

Fig. 10. *Screenshot of SVM active learning playlist generator. Song with a blue background is currently playing, and songs play down the list, with the user able to skip any song she does not want to hear. Green text indicates suggestions from playlist generator that have been accepted, red text indicates suggestions that have been rejected, and grey text indicates suggestions that have not yet been rated. This particular playlist was created by a user in the mood for music in the "rap" genre.*

Angularly Diverse Active Ranking: diverse active ranking using an angular diversity algorithm similar to that in Section V-B to choose a diverse set of instance from the lowest Clarity Index indexes for labeling by the user;

Information-Theoretic Diverse Active Ranking: diverse active ranking using an information-theoretic diversity algorithm similar to that in Section V-B to choose a diverse set of instance from the lowest Clarity Index indexes for labeling by the user.

Comparative testing among these approaches can be seen in Fig. 11. Each of the these bar plots [Fig. 11(a)–(c)] shows the relative percentage increase in precision for using information-theoretic diverse active ranking versus random active ranking, plain active ranking, and angularly diverse active ranking, respectively. The percentage increase is plotted with respect to increasing number of returned images, and the cluster of bars for each returned image size corresponds to performance during six rounds of feedback. (The absolute precision values for these experiments were between 0.70–0.85.)

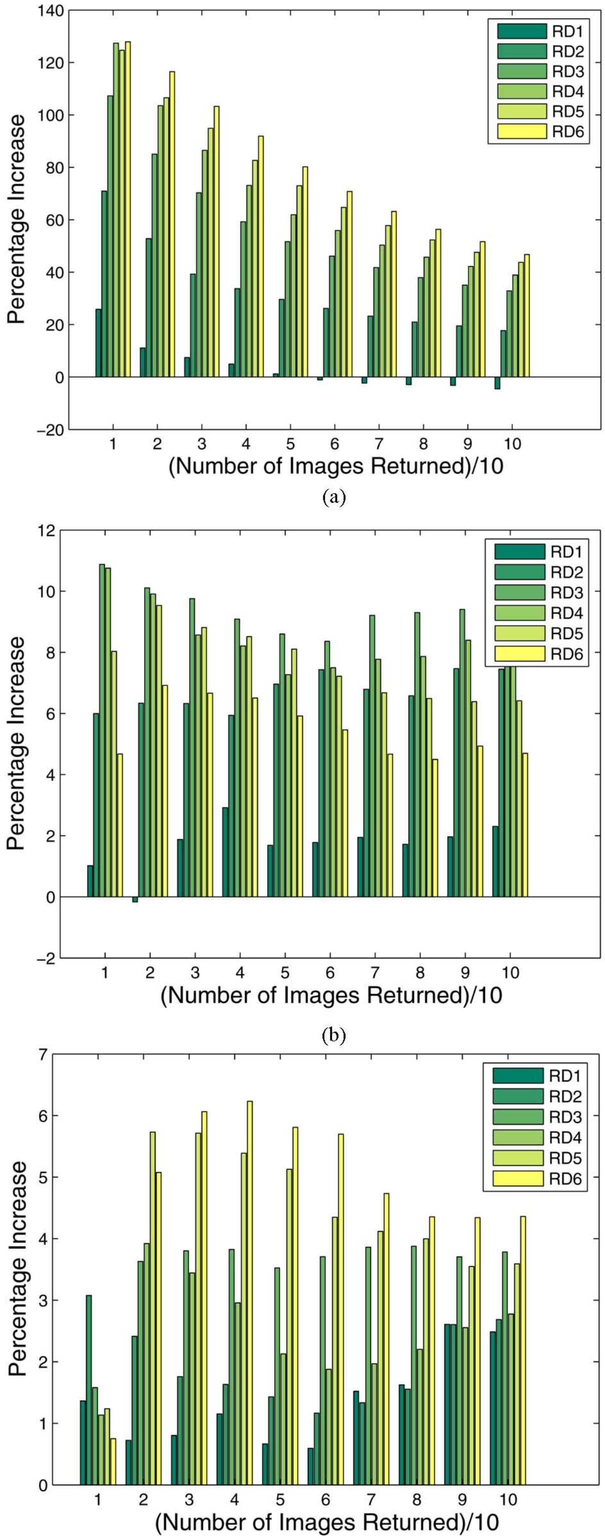
It was observed that information-theoretic diverse active ranking clearly outperformed both random active ranking and plain active ranking, by roughly 80%–100% and 8%–10%, respectively. Additionally, using the parameter-free information-theoretic diversity measure is on-par or better than the angular-diversity technique.

More interestingly, these results point toward an empirical property of diverse active ranking algorithms, namely their tendency towards improvement at lower rounds and lower numbers of returned images. This has practical significance in an information retrieval scenario. Users tend not to spend much time giving feedback or exploring many pages of returns to find what they are looking for. The low-end performance bias of this diverse active ranking system helps users find what they are looking for quickly and without scanning multiple pages of results. For more information, the interested reader is referred to [53].

VII. FUTURE OF ACTIVE LEARNING IN MULTIMEDIA INFORMATION RETRIEVAL

The future of interactive multimedia information retrieval systems should ideally exist online. In the past several years, we have seen a convergence-in-scale of internet infrastructure and content as well as human ability and effort. As a result, a vast amount of both data and human potential is waiting for the right problem and interactive algorithm to be applied to it. This, in turn, presents a wonderful opportunity for a re-imagining of the field of active learning for multimedia search and retrieval.

One will observe a steady trickle of works in this regard emerging from the research community. Recent work has sought to develop stand-alone interactive multimedia document search technologies [18], [63], organize a large image



(a)

(b)

(c)

Fig. 11. ***Comparison of percentage increase in precision obtained when using the information-theoretic diversity with active ranking (a) as compared to random active ranking, (b) plain active ranking, and (c) angularly diverse active ranking.***

collection of news images via person-clustering [67], collaboratively annotate large collections of images [68], [69], as well as look for automatic techniques to find iconic views of image queries for the online image sharing site Flickr [70].

The potential of such technologies lies not only in creating value for the user (as in the previous examples) but for the content-provider as well. Interactive multimedia information retrieval technologies can be used, for example, identifying duplication in text-indexed image and video databases thus providing more efficient indexing, as well as advertisement placement for multimedia documents.

There are a variety of reasons to focus on the eventual development of interactive technologies for multimedia in this space. The primary reason is that these real-world problems, though difficult and high-risk, have a high-reward potential. There is room for technologies that add great value to both the consumer and provider experience, and as such a successful application in this vein can stimulate funding for continued advances. Secondly, and more interesting from a research perspective, the vast amount of web data carries the potential to help with both generalization and discrimination problems. There is little work that looks at indexing and search with multimedia at such large scales.

There are many open questions in this regard: How do we adopt our interactive learning algorithms to work with such large scales? In addition, how will we resolve, from a fundamental research perspective, the large performance gain for even the simplest algorithms for all this data? Finally, when considering *interactive* technologies, how do we best utilize the effort of large numbers of users through either active (or passive) feedback? In addition to the traditional user-in-the-loop scenario, can we envision new paradigms that incorporate user-interaction *holistically* into the system? For example, can we use the implicit feedback within a social-networking structure to gain insights about the multimedia content it refers to?

The progress made within the last decade in interactive multimedia information retrieval given somewhat constrained application scenarios is very heartening. At this current juncture (lots of data and ready users), the field is standing at the base of a mountain of possibility. The opportunity exists to realize previously impossible-to-implement visions of automatic search, organization, and interaction with multimedia information. The challenges lie not only in determining what these new ideas will be, but also fostering interdisciplinary collaboration between industry, computer vision, data mining, user-interaction, and information retrieval researchers to help turn them into reality.

VIII. CONCLUSION

Aristotle has been credited for saying, “Learning is not child’s play, we cannot learn without pain.” Though the pain of the user can be considered a necessary evil in many interactive pattern recognition problems, the goal of active learning strategies since the time of Laplace has been to reduce this pain. Whether through careful experimental design, or more recently through thoughtful sample labeling strategies, active learning seeks to close the gap between human interpretation and machine understanding of real-world phenomena.

Much effort has been expended on investigating strategies for active learning that have applications in the multimedia information retrieval domain. A natural fit for these problems, recent work has shown the potential of active learning approaches can make real headway towards the automatic understanding of multimedia information.

As the first decade of the new century draws to a close, multimedia information retrieval is entering a new era. A unique confluence of media production, consumption, and economics has the potential to create new research and commercial opportunities for interactive technologies. In this space, active learning has the potential to be at the forefront of this technological movement, reducing the pain of learning for a brand new generation of interactive applications. ■

REFERENCES

[1] P. Graham. (2005, Nov.). *Web 2.0*. [Online]. Available: http://www.paulgraham.com/web20.html

[2] BBC News Inc. (2005, Jul.). *News Corp in $580 m Internet Buy*. [Online]. Available: http://www.news.bbc.co.uk/1/hi/business/4695495.stm

[3] News Corporation, Inc. (2003, Jul.). *Myspace: A Place for Friends*. [Online]. Available: http://www.myspace.com

[4] Facebook, Inc. (2004, Feb.). *Facebook—Welcome to Facebook!* [Online]. Available: http://www.facebook.com

[5] Ludicorp, Inc. (2004, Feb.). *Welcome to Flickr—Photo Sharing*. [Online]. Available: http://www.flickr.com

[6] BlueBridge Technologies Group. (2006, Feb.). *Zooomr: Experience the World Through Photos*. [Online]. Available: http://www.zooomer.com

[7] MSNBC.com News. (2006, Oct.). *Google buys YouTube for $1.65 Billion*. [Online]. Available: http://www.msnbc.msn.com/id/15196982/

[8] Google Blogoscoped. (2007, May). *New: Google Image Search Categories*. [Online]. Available: http://www.blogoscoped.com/archive/2007-05-28-n84.html

[9] ExaBlog. (2007, Apr.). *Exalead Becomes a ‘Facial Recognition Expert*.’ [Online]. Available: http://www.blog.exalead.com/2007/04/23/exalead-becomes-a-“facial-recognition-ex%pert”/

[10] Y. Rui and T. S. Huang, “Relevance feedback techniques in image retrieval,” in *Principles of Visual Information Retrieval*. London, U.K.: Springer-Verlag, 2001.

[11] Y. Rui, T. Huang, and S. F. Chang, “Image retrieval: Current techniques, promising directions and open issues,” *J. Visual Commun. Image Representation*, vol. 10, no. 4, pp. 39–62, Apr. 1999.

[12] C. G. M. Snoek and M. Worring, “Multimodal video indexing: A review of

the state-of-the-art," *Multimedia Tools Applic.*, vol. 25, 2005.

[13] X. S. Zhou, Y. Rui, and T. S. Huang, *Exploration of Visual Data*. New York: Kluwer, 2003.

[14] A. Dong and B. Bhanu, "Active concept learning for image retrieval in dynamic databases," in *Proc. IEEE Int. Conf. Computer Vision ICCV'03*, 2003.

[15] B. C. Ko and H. Byun, "Probabilistic neural networks supporting multi-class relevance feedback in region-based image retrieval," in *Proc. IAPR Int. Conf. Pattern Recognition ICPR'02*.

[16] J. P. Eakins, J. M. Boardman, and M. E. Graham, "Similarity retrieval of trademark images," *IEEE MultiMedia*, vol. 5, no. 2, pp. 53–63, 1998.

[17] N. Provos and P. Honeyman, *Detecting steganographic content on the internet*.

[18] C. G. M. Snoek and M. Worring, "Goalgle: A soccer video search engine," in *Proc. IEEE Int. Conf. Multimedia and Expo (ICME'03)*, Baltimore, MD, Jul. 2003.

[19] IBM Inc. (2006, Jul.). *IBM Multimedia Analysis and Retrievel System (Marvel)*. [Online]. Available: http://www.research.ibm.com/marvel

[20] Recording Industry Assoc. Amer. (2003, Dec.). *Press Room—Court/Legal Filings*. [Online]. Available: http://www.riaa.com/news/filings/default.asp

[21] Apple Computer, Inc. (2002, Dec.). *Itunes Music Store*. [Online]. Available: http://www.apple.com/itunes

[22] Microsoft, Inc. (2006, Nov.). *Zune.net—Home*. [Online]. Available: http://www.zune.net

[23] Pandora Media Inc. (2005, Dec.). *Pandora Internet Radio—Find New Music, Listen to Free Web Radio*. [Online]. Available: http://www.pandora.com

[24] M. Mandel, G. Poliner, and D. Ellis, "Support vector machine active learning for music retrieval," *Multimedia Systems Special Issue on Machine Learning Approaches to Multimedia Information Retrieval*, vol. 12, no. 1, pp. 3–13, 2006.

[25] J. Foote, "Content-based retrieval of music and audio," in *Multimedia Storage Archiving Systems II, Proc. SPIE*, 1997, pp. 138–147.

[26] M. Fink, M. Covell, and S. Baluja, "Social- and interactive-television applications based on real-time ambient-audio identification," in *Proc. EuroITV-06, 4th Eur. Interactive TV Conf.*, Athens, Greece, 2006.

[27] X. Zhou and T. S. Huang, "Small sample learning during multimedia retrieval using biasmap," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR'01)*, Kauai, HI, Dec. 2001.

[28] S. Mehrotra, K. Chakrabarti, M. Ortega, Y. Rui, and T. S. Huang, "Multimedia analysis and retrieval system (mars) project," in *Proc. Int. Workshop Multimedia Information Syst.*, 1997.

[29] H. K. Lee and S. I. Yoo, "A neural network-based image retrieval using nonlinear combination of heterogeneous features," *Int. J. Computational Intelligence Applic.*, vol. 1, no. 2, pp. 137–149, 2001.

[30] J. Edwards, R. White, and D. A. Forsyth, "Words and pictures in the news," in *Proc. NAACL Human Language Technology Conf. Workshop Learning Word Meaning From Non-Linguistic Data (NAACL'03)*, Edmonton, Canada, May 2003, pp. 6–13.

[31] M. Nakazato, C. Dagli, and T. S. Huang, "Evaluating group-based relevance feedback for content-based image retrieval," in *Proc. IEEE Int. Conf. Image Processing (ICIP'03)*, Barcelona, Spain, Sep. 2003.

[32] D. A. Cohn, Z. Ghahramani, and M. I. Jordan, "Active learning with statistical models," in *Advances in Neural Information Processing Systems*, vol. 7, G. Tesauro, D. Touretzky, and T. Leen, Eds. Cambridge, MA: MIT Press, 1995, pp. 705–712.

[33] N. Roy and A. McCallum, "Toward optimal active learning through Monte Carlo estimation of error reduction," in *Proc. Int. Conf. Machine Learning*, 2001.

[34] Y. Freund, H. S. Seung, E. Shamir, and N. Tishby, "Selective sampling using the query by committee algorithm," *Machine Learning*, vol. 28, no. 2–3, pp. 133–168, 1997.

[35] D. Angluin, "Queries and concept learning," *Machine Learning*, vol. 2, no. 4, pp. 319–342, 1988.

[36] E. T. Jaynes, "Bayesian methods: General background," in *Maximum Entropy and Bayesian Methods in Applied Statistics*, J. H. Justice, Ed. Cambridge, U.K.: Cambridge Univ. Press, 1986, pp. 1–25.

[37] D. V. Lindley, *On a Measure of the Information Provided by an Experiment,* vol. 27, pp. 986–1005, Dec. 1956.

[38] V. V. Federov, *Theory of Optimal Experiments*. New York: Academic, 1972.

[39] D. T. Davis and J. N. Hwang, "Attentional focus training by boundary region data selection," in *Proc. IEE Int. Joint Conf. Neural Networks IJCNN'92*, pp. 676–681.

[40] H. S. Seung, M. Opper, and H. Sompolinsky, "Query by committee," *Computational Learning Theory*, pp. 287–294, 1992.

[41] S. Tong and E. Chang, "Support vector machine active learning for image retrieval," in *Proc. IEEE Int. Conf. Computer Vision Pattern Recognition*, 2001.

[42] K. Goh, E. Y. Chang, and W.-C. Lai, "Concept-dependent multimodal active learning for image retrieval," in *Proc. ACM Int. Conf. Multimedia (MM'04)*, pp. 564–571.

[43] K. Brinker, "Active learning of label ranking functions," in *Proc. 21st Int. Conf. Machine Learning*, 2004, pp. 129–136.

[44] C. Dagli, S. Rajaram, and T. S. Huang, "Combining diversity-based active learning with discriminant analysis in image retrieval," in *Proc. IEEE Int. Conf. Information Technology Applications ICITA*, Sydney, Australia, Jul. 2005.

[45] H. T. Nguyen and A. Smeulders, "Everything gets better all the time, apart from the amount of data," in *Proc. ACM Conf. Image and Video Retrieval (CIVR'04)*, Dublin, U.K., 2004.

[46] Y. Song, X. S. Hua, L. R. Dai, and M. Wang, "Semi-automatic video annotation based on active learning with multiple complementary predictors," in *Proc. 7th ACM SIGMM Int. Workshop Multimedia Information Retrieval*, 2005, pp. 97–104.

[47] A. G. Hauptmann, J. Gao, R. Yan, Y. Qi, J. Yang, and H. D. Wactlar, "Automated analysis of nursing home observations," *IEEE Pervasive Computing*, vol. 3, no. 2, pp. 15–21, Apr.–Jun. 2004.

[48] B. Scholkopf and A. J. Smola, *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA: MIT Press, 2002.

[49] K. Porkaew and K. Chakrabarti, "Query refinement for multimedia similarity retrieval in mars," in *MULTIMEDIA '99: Proc. 7th ACM Int. Conf. Multimedia (Part 1)*, pp. 235–238.

[50] Y. Ishikawa, R. Subramanya, and C. Faloutsos, "Mindreader: Querying databases through multiple examples," in *Proc. 24th Int. Conf. Very Large Data Bases, VLDB*, 1998, pp. 218–227.

[51] D. H. Kim and C. W. Chung, "Qcluster: Relevance feedback using adaptive clustering for content-based image retrieval," in *SIGMOD'03: Proc. 2003 ACM SIGMOD Int. Conf. Management of Data*, pp. 599–610.

[52] X. S. Zhou, A. Garg, and T. S. Huang, "A discussion of nonlinear variants of biased discriminants for interactive image retrieval," in *Proc. ACM Int. Conf. Image and Video Retrieval (CIVR'04)*, Dublin, U.K., Jun. 2004.

[53] S. Rajaram, C. K. Dagli, and T. S. Huang, "Diverse active ranking for content-based search," in *Proc. IEEE Workshop Semantic Learning Applications in Multimedia (SLAM'07)*, Minneapolis, MN, Jun. 2007.

[54] Y. Freund, R. Iyer, R. E. Schapire, and Y. Singer, "An efficient boosting algorithm for combining preferences," in *Proc. ICML-98, 15th Int. Conf. Machine Learning*, J. W. Shavlik, Ed. San Francisco, CA: Morgan Kaufmann, 1998, pp. 170–178.

[55] S. Agarwal, T. Graepel, R. Herbrich, S. Har-Peled, and D. Roth, "Generalization bounds for the area under the ROC curve," *J. Machine Learning Res.*, vol. 6, pp. 393–425, 2005.

[56] C. Cortes and M. Mohri, *AUC Optimization Versus Error Rate Minimization*, 2004.

[57] T. M. Mitchell, "Generalization as search," *Artificial Intell.*, vol. 18, pp. 203–226, 1982.

[58] C. K. Dagli, S. Rajaram, and T. S. Huang, "Leveraging active learning for relevance feedback using an information theoretic diversity measure," in *Proc. ACM Conf. Image Video Retrieval CIVR'06*, Phoenix, AZ, Jul. 2006.

[59] C. K. Dagli, S. Rajaram, and T. S. Huang, "Utilizing information theoretic diversity for SVM active learning," in *Proc. IAPR Int. Conf. Pattern Recognition ICPR'06*, Hong Kong, China, Aug. 2006.

[60] P. Viola, N. N. Schraudolph, and T. J. Sejnowski, "Empirical entropy manipulation for real-world problems," in *Advances in Neural Information Processing Systems*, vol. 8, D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, Eds. Cambridge, MA: MIT Press, 1996, pp. 851–857.

[61] COREL Inc. (1998, Jan.). *Corel Stock Photo Images*. [Online]. Available: http://www.corel.com

[62] E. Y. Chang, S. Tong, K. Goh, and C. W. Chang, "Support vector machine concept-dependent active learning for image retrieval," *IEEE Trans. Multimedia*, 2005.

[63] VIMA Technologies Inc. (2003, Dec.). *Imagebeagle-Know What's on Your Computers*. [Online]. Available: http://www.imagebeagle.com/

[64] M. I. Mandel and D. P. W. Ellis, "Song-level features and support vector machines for music classification," in *Proc. 6th Int. Conf. Music Information Retrieval (ISMIR)*, J. D. Reis and G. A. Wiggins, Eds., Sep. 2005, pp. 594–599.

[65] D. Ellis, A. Berenzweig, and B. Whitman. (2005). *The 'Uspop2002' Pop Music Data Set*. [Online]. Available: http://www.labrosa.ee.columbia.edu/projects/musicsim/uspop2002.html

[66] All Media Guide, LLC. (2005). *Allmusic Guide*. [Online]. Available: http://www.allmusic.com

[67] T. Berg, A. C. Berg, J. Edwards *et al.*, "Names and faces in the news," in *Proc. IEEE Computer Vision Pattern Recognition (CVPR'04)*, Washington, DC, Jun. 2004, pp. 848–854.

[68] B. C. Russell, A. Torralba, K. Murphy, and W. T. Freeman, "Labelme: A database and web-based tool for image annotation," in *MIT AI Lab Memo AIM-2005-025*, Sep. 2005.

[69] Carnegie Mellon University. (2005, Dec.). *Peekaboom: Not Just Wasting Your Time*. . .. [Online]. Available: http://www.peekaboom.org

[70] T. Berg and D. A. Forsyth, "Automatic ranking of iconic images," Berkeley, CA, U.C. Berkeley Tech. Rep., Dec. 2006.

ABOUT THE AUTHORS

Thomas S. Huang (Fellow, IEEE) received the B.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan, China, and the M.S. and Sc.D. degrees in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge.

He was on the Faculty of the Department of Electrical Engineering, MIT, from 1963 to 1973. He was on the Faculty of the School of Electrical Engineering and Director of its Laboratory for Information and Signal Processing, Purdue University, from 1973 to 1980. In 1980, he joined the University of Illinois, Urbana-Champaign, where he is now the William L. Everitt Distinguished Professor of Electrical and Computer Engineering, and Research Professor at the Coordinated Science Laboratory, and Head of the Image Formation and Processing Group at the Beckman Institute for Advanced Science and Technology and Co-chair of the Institute's major research theme Human Computer Intelligent Interaction. His professional interests lie in the broad area of information technology, especially the transmission and processing of multidimensional signals. He has published 20 books and over 500 papers in network theory, digital filtering, image processing, and computer vision.

Dr. Huang was awarded the IEEE Third Millennium Medal, in 2000. Also in 2000, he received the Honda Lifetime Achievement Award for "contributions to motion analysis." In 2001, he received the IEEE Jack S. Kilby Medal. In 2002, he received the King-Sun Fu Prize, International Association of Pattern Recognition, and the Pan Wen-Yuan Outstanding Research Award. In 2005, he received the Okawa Prize. In 2006, he was named by IS&T and SPIE as the Electronic Imaging Scientist of the year. He is a Founding Editor of the *International Journal Computer Vision, Graphics, and Image Processing* and Editor of the Springer Series in Information Sciences, published by Springer Verlag. He is a member of the National Academy of Engineering, a foreign member of the Chinese Academies of Engineering and Sciences, and a Fellow of the International Association of Pattern Recognition, and the Optical Society of American. He has received a Guggenheim Fellowship, an A.V. Humboldt Foundation Senior U.S. Scientist Award, and a Fellowship from the Japan Association for the Promotion of Science. He received the IEEE Signal Processing Society's Technical Achievement Award, in 1987, and the Society Award in 1991.

Charlie K. Dagli received the B.S. degree in electrical and computer engineering from Boston University, Boston, MA, in 2001, and the M.S. degree in electrical engineering from the University of Illinois, Urbana-Champaign, in 2003, where he is currently pursuing the Ph.D. degree under the guidance of Prof. T. S. Huang.

His research interests are in applying interactive pattern recognition and information retrieval techniques to problems of multimedia search, retrieval, and clustering.

Mr. Dagli was the recipient of the Best Student Paper award at the 2006 International Conference on Image and Video Retrieval and the Best Student Paper Award in Pattern Recognition and Basic Technologies at the 2006 International Conference on Pattern Recognition.

Shyamsundar Rajaram received the B.S. degree in electrical engineering from the University of Madras, India, in 2000, the M.S. degree in electrical engineering from the University of Illinois, Chicago, in 2002, and the Ph.D. degree from the University of Illinois, Urbana-Champaign, under Prof. T. S. Huang.

He is currently a Senior Researcher in the Content-Analysis and Data Mining Group, Hewlett-Packard Laboratories, Palo Alto, CA. He has published several papers in the field of machine learning and its applications in signal processing, computer vision, information retrieval, and other domains.

Edward Y. Chang received the M.S. degree in computer science and the Ph.D. degree in electrical engineering from Stanford University, Palo Alto, CA, in 1994 and 1999, respectively.

He joined the Department of Electrical and Computer Engineering, University of California, Santa Barbara, in September 1999. He received his tenure in March 2003 and was promoted to Full Professor of electrical engineering in 2006. He is currently on leave from the University of California, heading Google/China research. Recent research contributions of his group include methods for learning image/video query concepts via active learning with kernel methods, formulating distance functions via dynamic associations and kernel alignment, managing and fusing distributed video-sensor data, categorizing and indexing high-dimensional image/video information, and speeding up support vector machines via parallel matrix factorization and indexing. His recent research activities are in the areas of machine learning, data mining, high-dimensional data indexing, and their applications to image databases, video surveillance, and Web mining.

Dr. Chang has served on several Association for Computing Machinery (ACM), IEEE, and Society for Industrial and Applied Mathematics (SIAM) conference program committees. He co-founded the annual ACM Video Sensor Network Workshop and has co-chaired it since 2003. In 2006, he co-chaired three international conferences: Multimedia Modeling (Beijing, China), SPIE/IS&T Multimedia Information Retrieval (San Jose, CA), and ACM Multimedia (Santa Barbara, CA). He serves as an Associate Editor for the IEEE Transactions on Knowledge and Data Engineering and *ACM Multimedia Systems Journal*. He is a recipient of the IBM Faculty Partnership Award and the NSF Career Award.

Michael I. Mandel received the B.S. degree in computer science from the Massachusetts Institute of Technology, Cambridge, and the M.S. degree in electrical engineering from Columbia University, in 2006. Currently, he is working toward the Ph.D. degree at Columbia University, New York.

He has published on music recommendation, music similarity, and sound source separation. His research uses machine learning to model sound perception and understanding.

Graham E. Poliner received the B.S. degree in electrical engineering from the Georgia Institute of Technology, Atlanta, in 2002, and the M.S. degree in electrical engineering from Columbia University, New York, in 2004. He is currently working toward the Ph.D. degree at Columbia University.

His research interests include the application of signal processing and machine learning techniques toward music information retrieval.

Daniel P. W. Ellis received the Ph.D. degree in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge.

He is an Associate Professor in the Electrical Engineering Department, Columbia University, New York. His Laboratory for Recognition and Organization of Speech and Audio (LabROSA) is concerned with all aspects of extracting high-level information from audio, including speech recognition, music description, and environmental sound processing. He also runs the AUDITORY e-mail list of 1700 worldwide researchers in perception and cognition of sound. He worked at MIT, where he was a Research Assistant at the Media Lab, and he spent several years as a Research Scientist at the International Computer Science Institute, Berkeley, CA.