ONTOLOGY-BASED ANNOTATION OF PAINTINGS WITH ARTISTIC CONCEPTS

MARCHENKO YELIZAVETA

(MSC. DONETSK STATE TECHNICAL UNIVERSITY, UKRAINE)

A THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN COMPUTER SCIENSE

SCHOOL OF COMPUTING

NATIONAL UNIVERSITY SINGAPORE

2007

Dedication

To my parents, Alla and Yevgen Marchenko

Acknowledgements

I wish to express my gratitude to everyone who contributed to this thesis. Specifically, I must single out my supervisor, Dr. Chua Tat-Seng, who gave his approval to this research topic and supported it throughout the years it took to bring it to fruition. I appreciate his vast knowledge of many research areas and his very patient assistance in helping me write many reports (i.e., reports, papers and this thesis), which occasionally made my eyes burn due to excessive red ink. I am also deeply grateful for his thoughtful and kind guidance during graduate training. Another person to whom I should express my deepest gratitude is Dr. Ramesh Jain, for his unceasing support of research ideas. His expertise, understanding and valuable advice added considerably to my graduate experience.

I would like to thank the members of my committee, Dr. Golam Ashraf and Dr. Leow Wee Kheng for the assistance they provided at all levels of the research project.

Very special thanks go out to Dr. Irina Aristarkhova, without whose motivation I would not have considered a graduate career. At the time, Dr. Aristarkhova was the one professor who truly made a difference in my life. It was under her tutelage that I changed focus and became interested in new media. She provided me with direction, technical support and became more of a mentor and friend, than a professor. It was through her persistence and kindness that I was encouraged to apply for graduate training. I doubt that I will ever be able to convey my appreciation fully, but I owe her my eternal gratitude.

Special thanks to my family for the love and understanding they provided me through my entire life. I wish I could name you all, for without your commitment I would not have finished this thesis. To my dad, Yevgen Marchenko, for his advice at times of critical need. To Alla and Ganna Marchenko, my loving and loyal supporters. My very special thanks, to my fiancée and best friend, Neil Leslie for his love, support and genuine ability to give and share happiness. Never underestimate the power of your encouragement.

I must also acknowledge Milanko Prvacki and LASALLE-SIA for the provision of the expert knowledge used in this study. Further appreciation goes out to Dr. Nikolai Ivanov for provision of the mathematical support for parts of this study. I would like to thank my friends in the Multimedia Lab, particularly Lekha Chairsorn and Huaxin Xu, for our philosophical debates, exchanges of skills, and venting of frustration during my graduate program.

To conclude, I would like to thank the National University of Singapore, Cyberarts Initiative, and School of Computing for their technical and financial support.

Contents

Summary

This thesis focuses on the automatic annotation of paintings with artistic concepts. To achieve accurate annotation we employ domain knowledge that organizes artistic concepts into the three-level ontology. This ontology supports two strategies for the concept disambiguation. First, more detailed artistic concepts serve as cues for the annotation of high-level semantic concepts. Second, the ontology relationships among high-level semantic concepts facilitate their disambiguation and serve to annotate the collection images in accordance to existing domain knowledge.

In this thesis we propose a framework that utilizes the three-level ontology of artistic concepts to perform annotation of paintings. We demonstrate that the use of domain knowledge in combination with low-level features yields superior results as compared to the use of only low-level features. The proposed framework performs successful annotation of a wide variety of high-level artistic concepts. This framework can be easily extended to annotate an even wider range of artistic concepts.

We propose two methods to facilitate the annotation of visual color, brushwork and application-level concepts respectively. For annotation of artistic color concepts, we develop a set of domain-specific features and combine them with inductive learning techniques. By testing various expert-provided queries, we demonstrate the satisfactory performance of the proposed method. For annotation of brushwork concepts, we develop a novel transductive inference approach that utilizes multiple classifiers to annotate brushwork concepts. We develop several variants of the proposed method and compare their performance with several baseline systems. The transductive inference approach is extended to facilitate annotation of application-level concepts such as artist names, periods of art and painting styles. Our experiments indicate that we could achieve over 85% of precision and recall for the annotation of artist and painting style concepts and over 95% for the annotation of art period concepts.

Lastly, we outline the major contributions of this thesis and list possible directions for future work.

List of Tables

List of Figures

Chapter 1

Introduction

Digital media progressively invades our everyday life. With the advent of the World Wide Web, large volumes of information are digitized. Imagery constitutes an important sub-domain of the digital media. Currently digital images are widely used in e-commerce, medical archives, military etc. Similarly, various art galleries and museum also digitize their collections. Primarily, digital scans of paintings introduce more interactivity for the virtual gallery visitors as well as they serve in anti-fakery analysis, preservation [Brown et al., 2001], educational and art historical uses [Hollink et al, 2003; Smeulders et al., 2002].

Large collections of digital scans require flexible and effective techniques to retrieve the necessary information. Current art retrieval systems mostly target large heterogeneous collections. Often these systems facilitate querying by image examples. They mostly employ low-level features as a basis for image representation [Chang, 1992; Lew et al., 2006]. A number of user studies demonstrated that low-level features have indirect relation to human interpretation of visual information, and consequently to user queries. Moreover, query by examples is ambiguous and it is difficult to formulate a precise query based on low-level features. This mismatch creates the so-called semantic gap and decreases the usability of the retrieval systems. In contrast, querying by semantic concepts or keywords is more natural to the end user. However, it requires complete annotation of the dataset with semantic concepts. At the moment, all paintings collections are annotated manually [Getty Research Institute, 2000].

Paintings domain has a number of distinctive characteristics. First, experts categorize paintings into a vast number of categories. They include objects and themes depicted (similarly to the general domain images) as well as various visual and high-level artistic descriptions [Brilliant, 1988; Greenberg et al., 1993; Hastings et al., 1995]. Second, visual attributes of paintings based on colors, brushwork and composition represent a vocabulary of visual-level concepts for analysis and description of masterpieces [Arnheim, 1954; Canaday, 1981; Lazzari, 1990]. While this vocabulary provides limited cues to the objects depicted, it serves as a major basis to characterize abstract and high-level descriptions such as artist name, painting style, period of art, culture etc. Thus, new techniques should be developed to facilitate the analysis and annotation of visual concepts. Due to these characteristics manual annotation of paintings is tedious and time consuming. Recently, statistical machine learning approaches have been proposed to perform automatic and semi-automatic annotation of paintings [Forsyth et al., 1997; Fung et al., 1999; Nigam et al., 2000; Lavrenko, 2003; Barnard et al., 2001 and 2003]. However, their performance is usually limited due to the semantic gap. Moreover, they often require large amount of labeled data to derive inferences of semantic concepts. These problems motivated our research to perform automatic annotation of paintings collections.

1. 1 Motivation

There are several factors that motivate our research:

First, there are large collections of paintings that require annotation. Usually they have limited or no annotations. In the paintings domain, artistic concepts offer an extensive vocabulary of concepts for navigation through paintings collections. For effective searching and browsing, annotation of these concepts is desirable. Figure 1.1 demonstrates an example of automatic paintings annotation.

Second, domain knowledge about paintings organizes these concepts into a hierarchical structure, where visual concepts reinforce high-level semantic concepts. This hierarchical organization serves to narrow the semantic gap between low-level features and high-level semantic concepts.

Third, manually labeled data for paintings is often difficult to gather. For example, manual annotation of brushwork classes requires extensive expertise. Hence, it is desirable to minimize the manually labeled data required for the learning of artistic concepts.

Fourth, effective auto-annotation techniques for the paintings domain are highly desirable. The goal is to develop methods for effective auto-annotation of both visual and high-level artistic concepts using domain knowledge and limited training sets.

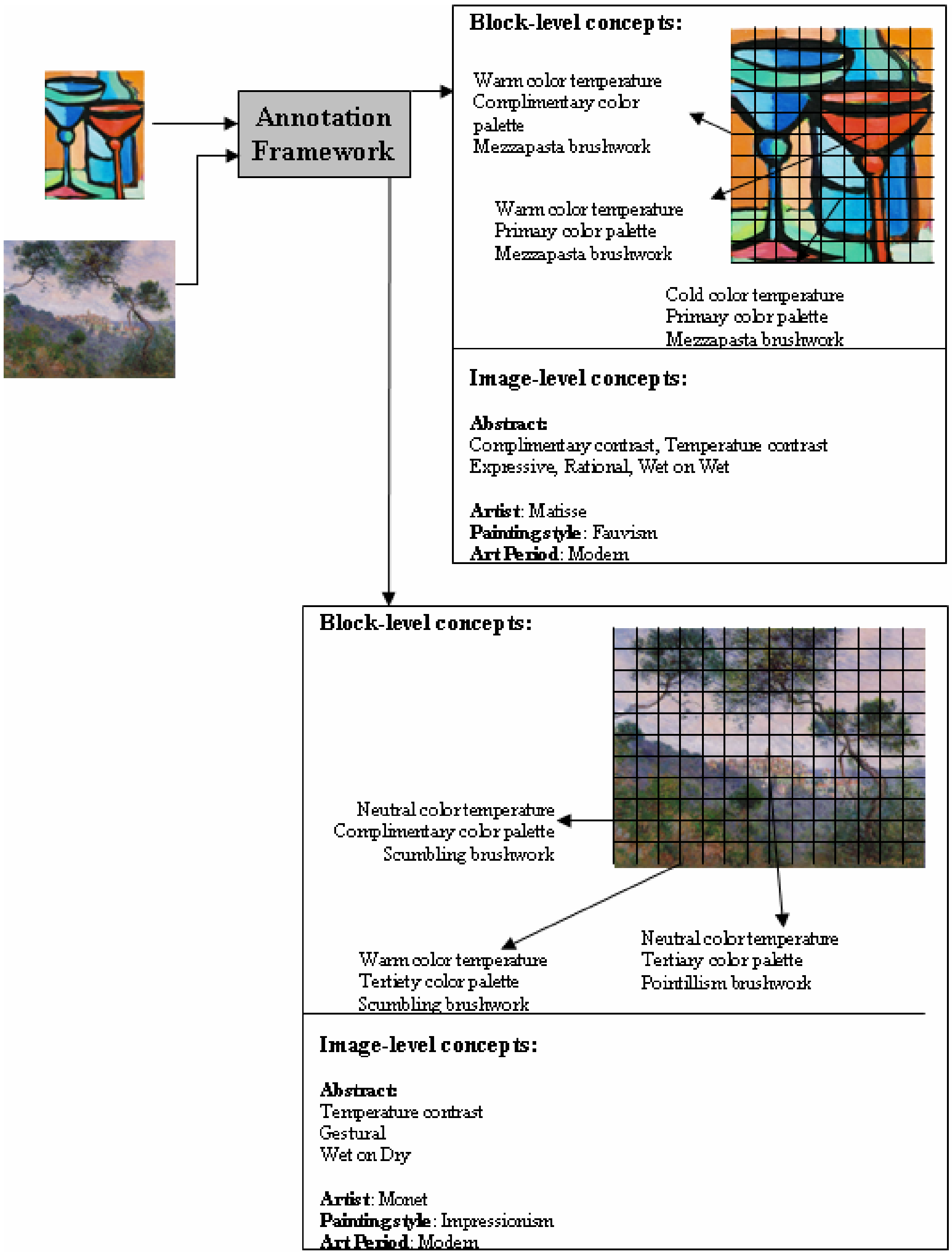


Figure 1. 1 Examples of automatic paintings annotation

1. 2 Our approach

In this dissertation, we propose a flexible framework that performs the annotation of paintings with artistic concepts using domain knowledge. This framework follows the hierarchical learning paradigm that mimics human cognition and reinforces hierarchical organization of artistic concepts.

Visual concepts describe image regions, while high-level semantic concepts usually describe the whole image. In accordance to hierarchical learning, we first assign visual-level concepts to the image region based on low-level features. Next, we combine low-level features and visual-level concepts to generate annotations of regions with respect to high-level concepts. Lastly, using the ontological relationships among high-level concepts we integrate region-based information and disambiguate these concepts to represent the whole image.

Figure 1.2 demonstrates relationship between the ontology of artistic concepts and the proposed framework.

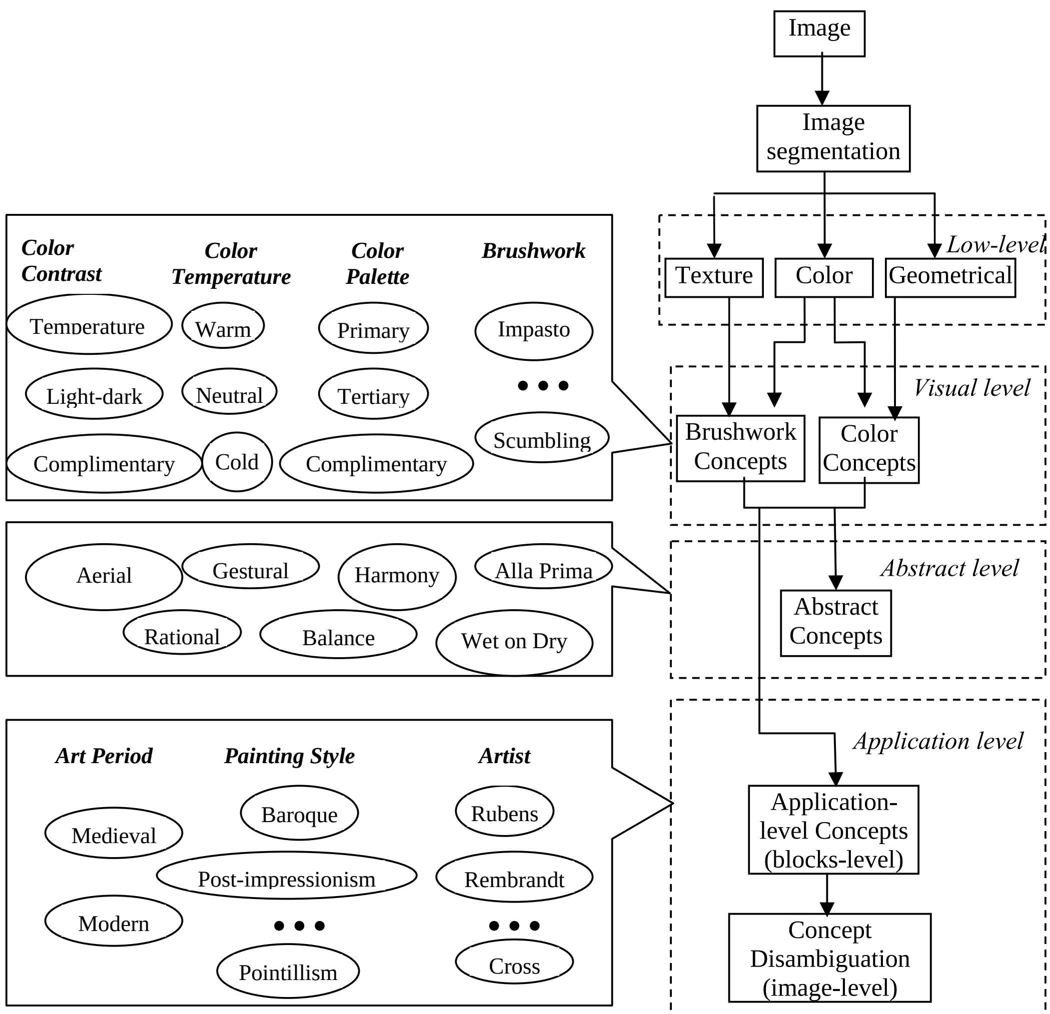


Figure 1. 2 Annotation of the ontology concepts within the proposed framework

This figure demonstrates how various levels of ontology correspondence to the hierarchical annotation process of the proposed framework. This framework incorporates domain ontology of artistic concepts that facilitates concept disambiguation and has a number of advantages for navigation and retrieval. The framework performs inference using different types of learners, both supervised and semi-supervised. This facilitates inferencing of the concepts that have limited amount of the labeled data. Overall, the proposed framework implements a range of methods for the annotation of visual-level color, brushwork as well as abstract and high-level semantic concepts.

Figure 1.3 demonstrates how these methods combine within the overall framework for paintings annotation. These methods include:

1. Fully supervised annotation of visual-level color concepts. To perform annotation, we employ the artistic color theory of Itten [1961]. This theory offers a mapping between color hues and visual-level color concepts. Our method extends existing works in several directions. First, for effective representation of image image, we extract domain-specific color features that represent the distribution of artistic concepts within a region. In our work we experiments with two types of image regions: a) color/texture blobs generated using image segmentation techniques; and 2) fixed-sized blocks. Second, we demonstrate that using visual-level concepts and their ontological relationships the proposed method facilitates the annotation of abstract artistic color concepts without additional training. Specifically, we employ the artistic color sphere and fully supervised probabilistic SVM classifier.
2. Semi-supervised annotation of brushwork patterns. To facilitate effective annotation of these complex patterns, we adopt the serial multi-expert approach, where sequentially arranged experts (learners) perform step-wise disambiguation of the target concepts based on a decision hierarchy. The decision hierarchy encodes relationships among classes, thus iteratively splitting a dataset into sub-classes until the leaf nodes with the target concepts are reached. Due to its modularity, this approach facilitates feature selection and model selection for each node of the decision tree. We combine this approach with semi-supervised learning methods to address the problem of limited labeled datasets. Using this method, we investigate: a) one-step annotation of brushwork classes and step-wise disambiguation using multiple experts; b) manual and automatic selection of low-level features and parameters of the semi-supervised learning methods and the use of distance-based and probabilistic semi-supervised learning methods. We aim to demonstrate that the resulting transductive inference using multiple experts is effective for the annotation of complex brushwork patterns and that the proposed methods for automatic feature

and parameter selection technique is comparable to the manually assigned features.

3. Annotation scheme for labeling high-level semantic concepts. This scheme includes two major steps: a) the annotation of image regions with high-level semantic concepts and b) the integration of the generated concepts to annotate the whole image. For step (a) we employ the semi-supervised techniques developed for brushwork annotation. In this step we exploit the fact that visual-level concepts serve as cues for annotation of high-level concepts.

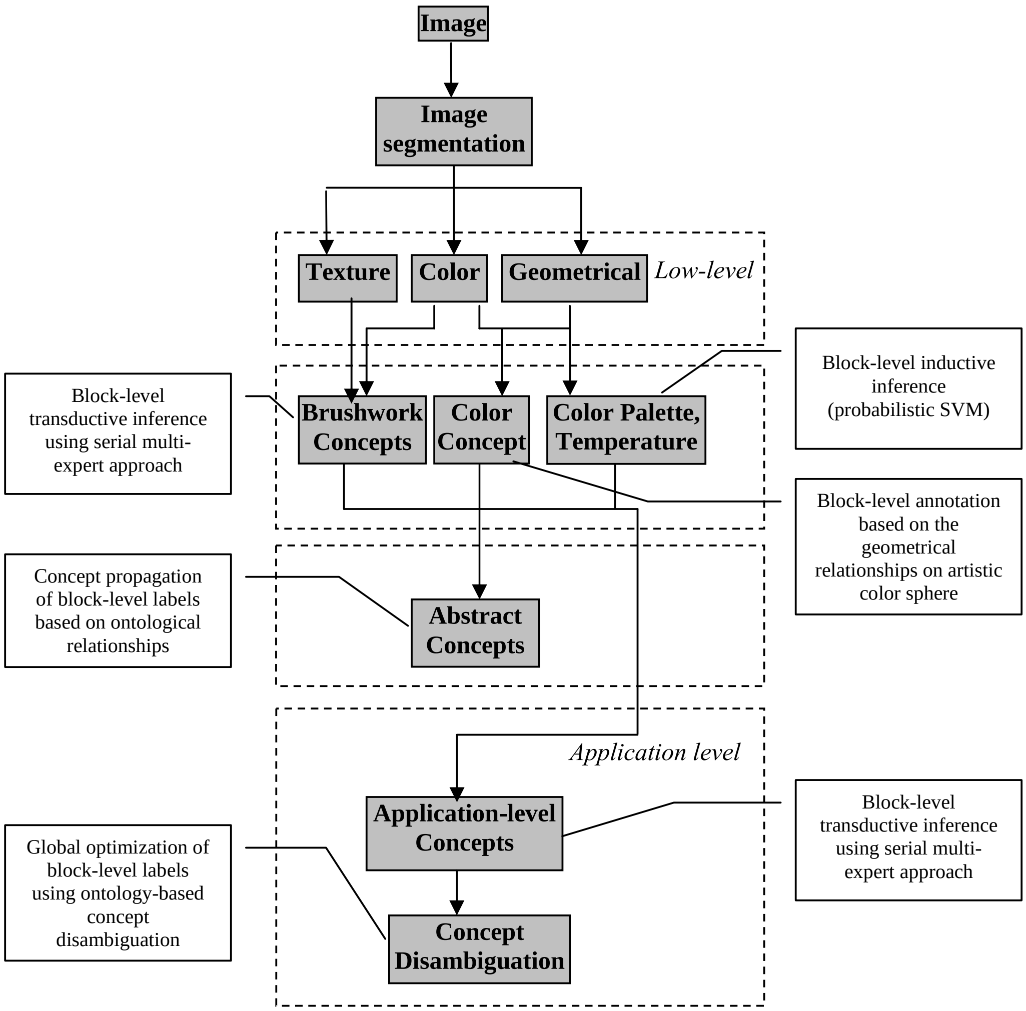


Figure 1. 3 High-level scheme of the proposed framework

We thus utilize the visual-level concepts as meta-level information and employ the transductive inference and multiple experts to label the whole image with high-level artistic concepts such as the artist name, painting style and art period. We aim to demonstrate: a) the importance of meta-level information in the annotation process; b) the effectiveness of multiple experts approach as compared to one-step inference approach and c) the effectiveness of the proposed method to generate satisfactory performance using limited training set. Next, using the generated labels, we further exploit the ontological relationships among high-level concepts to disambiguate concepts. We aim to demonstrate that ontological relationships are efficient as compared to the use of automatically generated results for the concept disambiguation.

1. 3 Contributions

In this thesis we make the following contributions:

1. We propose a novel framework for the annotation of paintings with artistic concepts using domain ontology. This ontology includes visual concepts and high-level concepts and relationships among them. This framework employs visual-level concepts as meta-level information and facilitates concept disambiguation based on the ontological relationship.
2. We propose and implement the method for annotation of visual color concepts that combines domain knowledge and machine learning techniques.
3. We propose and implement a transductive inference method for the annotation of brushwork visual concepts. This method utilizes multiple expert approaches that facilitates disambiguation of patterns and performs automatic selection of features and model parameters.
4. We extend the proposed transductive inference approach to perform the annotation of high-level concepts and their disambiguation based on ontological relationships.

1. 4 Thesis Overview

The dissertation is organized as follows:

Chapter 2 discusses the problem of automatic image annotation. It motivates the need for the machine learning approach and discusses the measures for performance evaluation.

Chapter 3 reviews the state-of-the-art approaches to image annotation and retrieval. It discusses the existing ontologies for manual annotation, the query by example and query by keyword paradigms. We further discuss semi-supervised and supervised learning approaches and ontology-based annotation.

Chapter 4 discusses the domain-specific knowledge used in our study. It presents a three-level organization of artistic concepts, where visual-level concepts reinforce abstract-level and application-level concepts. These concepts offer an extensive vocabulary for annotation.

Chapter 5 presents the proposed framework for the annotation of paintings with artistic concepts. This learning framework exploits domain specific knowledge in order to narrow down the semantic gap. It implements hierarchical learning, where the system first annotates image region, and then uses the region-based annotations to infer image-level labels.

In Chapter 6, we propose and implement an approach for supervised annotation of paintings with visual-level color concepts. This approach employs artistic theory to extract domain-specific features and annotate paintings.

In Chapter 7, we propose and implement a semi-supervised transductive approach to annotation of paintings with brushwork classes. This approach adopts multiple expert paradigm that facilitates step-wise disambiguation of the target concepts. We compare several variations of the proposed method based on different semi-supervised techniques and feature selection methods.

In Chapter 8, we employ the semi-supervised transductive method proposed in Chapter 7 to annotate image with semantic concepts. Using this method, we demonstrate that the use of visual-level artistic concepts is beneficial to the annotation of high-level concepts. We also propose a concept disambiguation method that utilizes ontological relationships among concepts.

Finally, Chapter 9 concludes the thesis with a discussion of future research.

Chapter 2

Automatic Annotation of Images

2. 1 Manual and Automated Annotation of Images in Paintings Domain

Image is a complex medium. As discussed in [Panofsky, 1962], there are at least three aspects that influence image interpretation. First, image can be “of” and “about” something. For example, an image is “of” a woman and a child and “about” immaculacy. Second, image contains, simultaneously, generic and specific information. The user might treat the object depicted in the image as the representation of this particular object (image of Titanic) or general concept of this object (image of Titanic as an example of a ship). Third, image can be broadly classified as being “of” or “about” time, space, activities and objects. Complexity of visual information introduces difficulties in the annotation process and naturally leads to the subjectivity of annotation.

In an attempt to embrace and standardize all possible interpretations of an image, researchers developed concept ontologies that serve for manual annotation. To describe paintings, human experts often use arts-oriented ontologies that include artistic and general concepts, which describe and characterize an image at various levels of detail. This includes visual characteristics of paintings as well as description of its objects, mood, theme etc. Majority of manual annotations serve for cataloguing and preservation purposes. The list of established ontologies for the description of visual documents and historical materials includes:

- ICONCLASS [Waal, 1985],
- Art and Architecture Thesaurus (AAT) [Getty Research Institute, 2000],
- United List of Artist Names (ULAN) [Getty Research Institute, 2000], and
- Thesaurus for Graphic Materials and Metadata (TGM) [Library of Congress, 2000].

These external ontologies represent a complex tool for manual annotation. Each of the ontologies includes a vast number of terms that require extensive knowledge of the respective domain from the annotators. In an attempt to assist in the annotation process, various researchers [Hollink et al., 2003, Hyvönen et al., 2003; Smeulders et al., 2002] developed ontology-based tools for annotation. However, even with these ontology-based tools, the human effort required for annotation is still substantial. To eliminate these efforts, a fully automated annotation system is desired. The purpose of such an annotation system is to automatically assign the appropriate concept labels to each image. The automatic annotation system analyzes an image using multiple concept learners and assigns multiple concepts that represent the content of an image. Semantic annotations of paintings can be used for the following purposes:

- Image retrieval using queries such as ‘paintings by Cezanne’, ‘paintings with warm colors on top’. Optionally the system may facilitate relevance feedback to utilize the user in the retrieval process.
- Ontology-based navigation of image collections – using ontology to provide context for navigation and querying of collections.
- Integration of image collections – ontology-based semantic annotations facilitate unified access to collections of various museums.
- Combining automatically annotated concepts with domain-specific knowledge serves to automatically compose a summary for each painting.

However, automatic annotation of paintings with semantic concepts is a challenging task for several reasons:

- The limited representational power of color and texture low-level features. For example, images with the same low-level features may have different contents. Similarly, an image under different lightning conditions is represented by different color feature vectors.
- Due to such reasons as light intensity, occlusions etc, the image segmentation task is difficult and its result is unstable. Thus, the image regions often do not correspond to meaningful objects, making the semantic annotations based on such regions incomplete or erroneous.
- High-level concepts may have a variety of visual representations and, thus, various values of low-level features.
- Automatic annotation does not incorporate relationships among concepts such as the synonyms.

2. 2 Machine Learning for Automated Annotation

In general there exist two approaches to problem solving: *knowledge engineering* and *machine learning*. In the knowledge engineering approach, a program aims to solve the problem directly using a set of rules. Determining a specific set of rules that applies to all kinds of images is a very difficult task.

The machine learning approach provides an indirect approach, wherein the system learns how to solve the problem of interest. As discussed in Mitchell [1997], machine learning denotes the acquiring of general concepts based on specific training samples. For concept learning task, machine learning aims to find an approximation of an *unknown target* function

$$\Phi: \{I, C\} \rightarrow \{T, F\} \tag{2. 1}$$

where I denotes a set of images (documents) that are members or non-members of concept of interest C. *T*he target function Φ in Equation 2.1 represents the classification an image $I_i \in I$ as whether is should be assigned to concept C and value F is the decision not to assign an image $I_i \in I$ to concept C. Φ describes how images I ought to be classified and, in short, assigns $I_i \in I$ to C. The approximation function

$$\Phi': \{I, C\} \rightarrow \{T, F\} \tag{2. 2}$$

is called *a classifier* and, ideally, should closely match Φ. *The c*lassifier stores parameters of approximation function or *hypothesis* in the knowledge base *KB*. This knowledge base is further applied to solve the previously unseen problems. This approach has one important assumption that unseen samples come from the same distribution as the samples used for training.

We employ the machine learning approach in our framework due to several reasons. First, it avoids the need to collect, organize and resolve large amounts of incomplete and conflicting human knowledge. Second, the use of machine learning makes the system very flexible: we can easily re-train the system with new training sets or to handle the new set of semantic concepts.

2. 3 Inductive and Transductive Learning

Machine learning largely relies on Statistical Learning Theory and its major concepts such as *induction*, *deduction*, and *transduction*. In classical philosophy, deduction describes the movement from general to particular, while induction denotes the movement from particular to general. Figure 2.1 depicts relationships between these learning concepts as discussed by Vapnik [1995]. Induction derives the *unknown target function* from given data, while deduction derives the values of the *given function* for points of interest.

The classical scheme [Vapnik; 1995] suggests that the derivation of the values of the target function for the points of interest proceeds in two steps: first using the inductive step, and then using the deductive step. The inductive inference for concept learning can be formalized using Formulae 2.2. The version space *Φ'* represents the subset of hypothesis in the hypothesis set *H* that are consistent with the training set *I*. Intuitive interpretation of the inductive inference formulation assumes the training set, where each training sample has pre-assigned values (or *label*) *T* or F that denote whether the current samples belongs to class *C*. An algorithm that learns from only labeled samples is called a *supervised* learner.

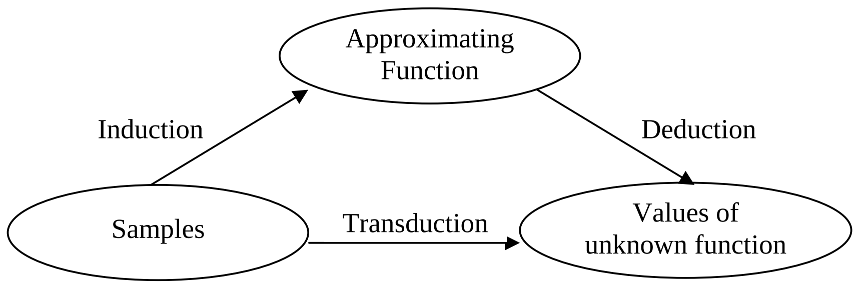


Figure 2.1. Types of Inference (by courtesy of Vapnik [1995])

As pointed out by Vapnik [1998] in many realistic situations one actually faces an easier problem, where one is given a training set of labeled examples, together with an unlabeled set of points which needs to be labeled. Such a type of inference is called *transductive* inference and denotes moving from particular to particular. In this transductive setting, one is not interested in inferring a general rule, but rather only in labeling this unlabeled set as accurately as possible. Using this type of inference, we derive the values of the unknown target function for the given data. One solution is of course to infer a rule as in the inductive setting, and then use it to label the required points. However, as argued by Vapnik [1982, 1998], it makes little sense to solve what appears to be an easier problem by `reducing' it to a more difficult one. While there are currently no formal results stating that transduction is indeed easier than induction, it is plausible that the relevant information carried by the test points can be incorporated into an algorithm, potentially leading to superior performance. Since a transductive learner facilitates inference based on both labeled and unlabelled samples, this type of setting assumes a *semi-supervised* learner. Similarly, an *unsupervised learner* is trained using solely unlabelled training samples. Various distance-based clustering techniques such as K-means serve as examples of unsupervised learners. They cluster the unlabelled samples based on their distances to the cluster centers.

We demonstrate the generic framework for supervised and semi-supervised learning in Figure 2.2. Both frameworks are very similar except that the semi-supervised learner utilizes different learning strategies as compared to the supervised learner. The raw data (includes scans of paintings in our case) are preprocessed to extract features for adequate data representation. In the training mode, as outlined by spotted-line box, the teacher (human expert) assigns the concepts to each training sample. Such assignment gives rise to the term *supervision*. Under semi-supervised paradigm, the learner composes the training set using both labeled and unlabelled samples available. As shown in Figure 2.2, the predictor utilizes the resulting knowledge to generate labels for previously unseen samples. In general, labeled samples are divided into training and testing sets. In our work, we utilize 315 and 735 images for training and testing respectively. These sets are often used to test the ability of the learner to construct an accurate and generalized knowledge base.

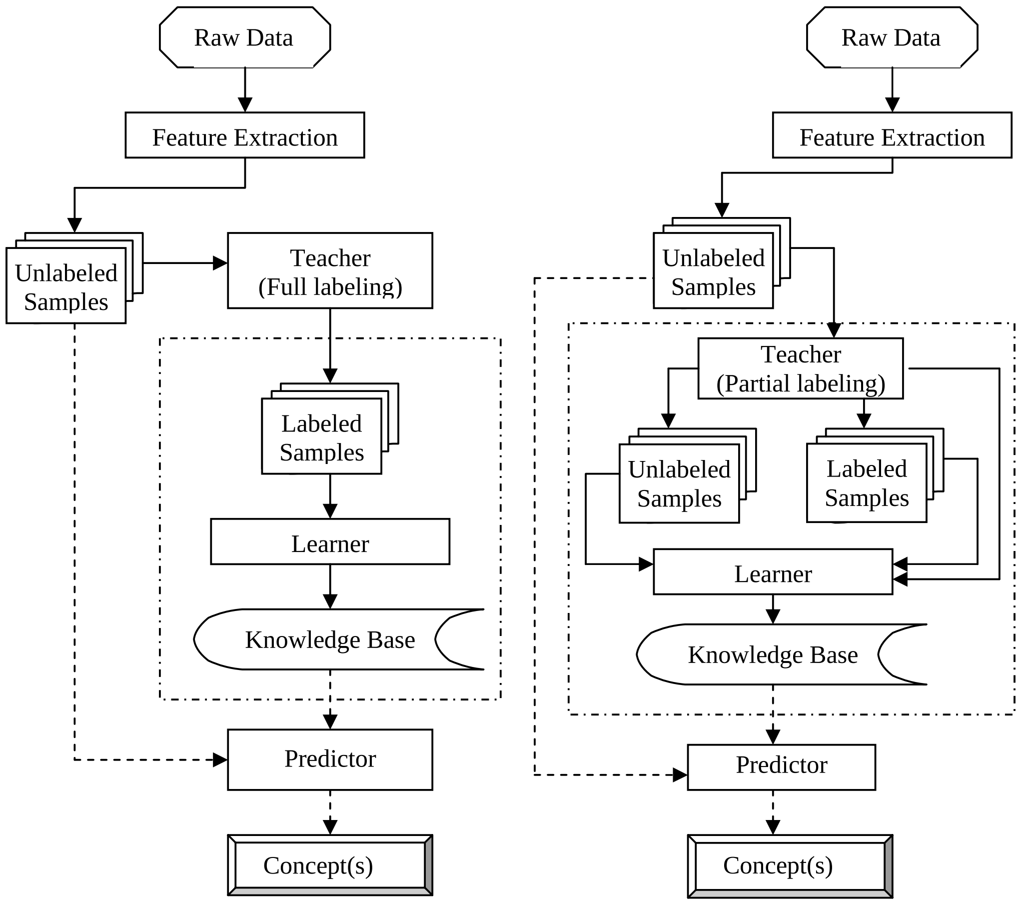


Figure 2. 2. Frameworks for supervised and semi-supervised learning

2. 4 Drawbacks of Machine Learning for Image Annotation

While numerous works demonstrated satisfactory performance of machine learning methods, it is still a challenging task for several reasons:

1. Mapping

   There is no clear mapping from a set of visual features to its semantic concepts. First, semantically different and visually similar objects/regions may have similar representation in terms of visual features. For example, a region of blue color may depict sky, water, blue wall etc. Similarly, in the paintings domain a region of coarse directed texture may represent brushwork technique of Cezanne, van Gogh or Seurat. Next, lightning conditions, occlusions and other factors change visual appearance of objects. Lastly, semantics of a regions indirectly relates to the semantic of the overall image. So given that we are able to capture semantic labels of an image we might not be able to capture the semantics of the overall image.

2. The curse of dimensionality

   The fundamental reason for this phenomenon is that high-dimensional functions have the potential to be much more complicated as compared to low-dimensional ones, and these complications are harder to discern [Duda et al., 2000]. The system requires a large number of samples to perform training in high-dimensional feature space, which in turn poses the need for substantial human effort for annotation. In general, the relationship between required samples and feature dimensionality is exponential, which restricts the application of machine-learning methods.

3. Feature irrelevance

   The majority of learners utilize all features available whether or not these features are relevant to the target concept, except for the rule-based and decision-tree approaches. Due to this, samples with similar relevant features might be far from each other. Thus, the similarity metrics based on the full feature space might be misleading since the distance between neighbors is likely to be dominated by the large number of irrelevant features. This problem is evident in paintings domain, where brushwork patterns exhibit a large variety of properties that requires a large number of low-level features.

4. Label noise

   Label noise refers to the fact that the labels assigned to the samples by the human annotator may contain errors. Annotation of image with wrong labels may be due to: (a) variations in human expert knowledge, (b) unreliable image segmentation and (c) image quality.

5. Domain knowledge and Concept relationships

   Traditional machine learning approaches are not aware of the relationships among concepts and concept granularity. This property of the machine learning approach contrasts with the human ability to conceptualize the world. For example, in the paintings domain concepts of different artist names should not appear within the same painting. Lack of such so-called *domain-specific* knowledge about relationships among concepts leads to the decreased accuracy of the machine learning systems.

2. 5 Performance Measurement

Since automatic annotation system is a natural base for information retrieval systems, there are two major approaches for its evaluation. First, we evaluate such a system using performance measures for the information retrieval system. Second, we utilize measures for performance evaluation of classifiers. The choice of the measures often depends on the characteristics of data collection, user needs etc. In this thesis we employ a variety of measures for the evaluation of our proposed framework.

2. 5. 1 Contingency Table

Contingency table is widely used for the evaluation of both classification and information retrieval tasks. In the context of classification task, contingency table demonstrates the distribution of classifier predictions into two or more categories. It is also known as confusion matrix. Table 2.1 demonstrates the 2x2 contingency table used for performance evaluation of binary classifiers or, in other words, classifiers that predict whether a sample belongs to a category or not.

	Actual Labels		
Predicted Labels		Negative	Positive
	Negative	TN	FN
	Positive	FP	TP

Table 2. 1 Contingency Table of 2x2 size

In context of an image annotation system, a sample denotes a unit of analysis (image or region, for example) and a category refers to a concept. The term “Positive” denotes that the samples belong to the category of interest and “Negative” that they do not belong to this category. Since we have information about true data labels and predicted data labels, the contingency table classifies samples into: False Positive (FP) if it predicts negative samples to be positive, False Negative (FN) if it predicts that samples are negative while they are actually positive, True Negative (TN) and True Positive (TP) if the system predicts the label of samples correctly. Hence, with this notation the number of correctly predicted samples is TP+TN, while prediction over all samples is equal to TP+TN+FP+FN.

To ease comparison of the tables, several performance measures have been developed based on the four values of the contingency table. Transforming four values into a single value usually causes some loss of information, due to which some measures are more preferable than others [Liere, 1999]. The following evaluation measures are widely used:

1. Sensitivity

   Sensitivity denotes the ratio of true positive predictions to the number of positive instances in the test set:

$$sensitivity = \frac{TP}{TP + FN}, \quad if \;\; TP + FN = 0, then \; sensitivity = 0 \tag{2. 3}$$

2. Specificity

Specificity denotes the ratio of true negative predictions to the number of negative instances in the testing set.

$$specificity = \frac{TN}{TN + FP}, \quad if \;\; TN + FP = 0, then \; specificity = 0 \tag{2. 4}$$

3. Accuracy

Accuracy measures the ability of the system to correctly predict label of samples. It is defined as the ratio between the number of correctly identified samples and the size of testing set:

$$\begin{aligned} &accuracy = \frac{TP + TN}{TP + FN + TN + FP} \\ &if \;\; TP + FN + FP + TN = 0, then \; accuracy = 0 \end{aligned} \tag{2. 5}$$

4. Precision and Recall

These two measures are commonly used for evaluation of information retrieval tasks. They represent the system evaluation in contrast to the user-based evaluation. The system evaluation is done in laboratory and, thus, is comparatively cheap. It was first performed over four decades ago by Cranfield [Cleverdon et al., 1966] and since then became a dominant IR model for such evaluation efforts as Text REtrieval Conference [Voorhees et al., 2006]. Precision characterizes the ability of the system to predict positive samples that are actually positive. It is defined as the ratio between the number of correctly identified samples and the number of totally identified positive samples:

$$precision = \frac{TP}{TP + FP}, \quad if \;\; TP + FP = 0, then \; precision = 0 \tag{2. 6}$$

Recall measures the ability of system to identify positive samples in the dataset. It is defined as a ratio between the number of true positive samples and the total number of positive samples in dataset:

$$recall = \frac{TP}{TP + FN}, \quad if \;\; TP + FN = 0, then \; recall = 0 \tag{2. 7}$$

During actual testing, the classification and retrieval system usually exhibits tradeoff between recall and precision.

2. 5. 2 Practical Performance Measures

In Section 2.5.1 we have discussed several widely used performance measures for evaluation of classification and retrieval systems. However, in practical applications, these performance measures have some changes. “True” and “false” sample labels are changed to the concepts of *relevance*. Thus, equations become:

$$precision = \frac{retrieved\ relevant\ samples}{retrieved\ samples}$$
$$recall = \frac{retrieved\ relevant\ samples}{relevant\ samples} \quad (2.\ 8)$$

Due to the fact that the degree of relevance is based on the user point of view, it introduces subjectivity to the evaluation of the system. In our task the order of retrieved samples might have importance. To evaluate the performance of the system we employ Mean Average Precision (MAP) metrics. Thses metrics favor highly ranked relevant items. To calculate average precision, we measure precision after each relevant document in a collection is retrieved. To calculate MAP we take the mean of average precision across all categories.

In actual practice, the classification systems exhibit precision-recall tradeoff. In comparing two systems, one always favors the one having higher precision and recall. To incorporate both recall and precision into a single value, [Lewis et al., 1994] proposed F_b measure. This measure is a function of recall, precision and a positive constant b, which represents the importance ratio of recall to precision:

$$F_b = \frac{(b^2+1)\times precision \times recall}{(precision + recall)},$$
$$if\ \ precision = 0\ and\ recall = 0,\ then\ F_b = 0 \quad (2.\ 9)$$

In our experiments, we give equal importance of recall and precision *(b=1)* to evaluate the proposed system.

In order to understand the experimental results better, we calculate precision, recall and F_1 measure using micro- and macro- averaging. Using macro-averaging, we calculate these measures for each category and then average. Using micro-averaging, we calculate them over all decisions. The two procedures bias the results differently - micro-averaging tends to over-emphasize the performance on the largest categories, while macro-averaging over-emphasizes the performance on the smallest. The analysis of these two measures gives insights to the distribution of data across categories.

In this chapter we discussed why auto-annotation of images is useful for annotation of artistic images and, in particular, paintings. We also introduced existing paradigms for machine learning and presented widely used evaluation measures. In the next chapter, we present the state-of-the-art works information retrieval and statistical learning systems and provide a basis for a framework for automatic annotation of paintings.

Chapter 3

Overview of Existing Work for Paintings Annotation

In this chapter, we focus on the existing studies on annotation and retrieval task for general images and, in particular, paintings. We then discuss existing problems and some strategies to overcome them.

3. 1 Existing Ontologies for Paintings Annotation

We start our discussion with existing arts-oriented ontologies that are widely used for the cataloguing and description of arts objects. The list of established ontologies for the description of visual documents and historical materials includes:

- ICONCLASS [Waal, 1985]
- Thesaurus for Graphic Materials and Metadata (TGM) [Library of Congress, 2000]
- Art and Architecture Thesaurus (AAT) [Getty, 2000]
- United List of Artist Names (ULAN) [Getty, 2000]

All these tools include a fixed vocabulary of the artistic concepts organized into a hierarchy. However, they differ in their scope of terms, level of details and applicability to arts collections.

The ICONCLASS ontology covers early and medieval art collections, in which theme, historical and religious aspects represent important concepts for description. It divides iconography into the following categories:

- Religion and Magic, Nature,
- Human Being and Man in General,
- Society, Civilization, and Culture,
- Abstract Ideas and Concepts,
- History,
- Bible,
- Literature,
- Classical Mythology

- Ancient History.

Clearly, this ontology of concepts maintains the traditional coherence of content with biblical, classical, historical or literary sources and is mostly useful for annotation of medieval arts collections.

The TGM ontology is meant for wider range of arts objects and collections. It contains the following facets at the highest level of the concept hierarchy:

- Geography
- Nationality
- Ethnic Group
- Racial Group
- Religion
- People
- Form and Genre
- Physical Characteristics.

In contrast to the TGM ontology, Art and Architecture Thesaurus (AAT) serve to describe visual artistic documents, for example, paintings, frescos, mosaic. Due to this, we extensively employ this ontology and its concepts in our framework. AAT includes 125,000 concepts organized under the following categories at the highest-level:

- Associated and Abstract Attributes,
- Physical Attributes,
- Styles and Periods,
- Agents,
- Activities,
- Materials and Objects.

The category of Associated and Abstract Attributes includes a variety of non-visual terms reflecting the content of painting. For example, it includes perceptual effects that are induced by the use of specific painting techniques. For example, it is widely accepted that the use of contrasting colors is regarded as expressive in the western fine arts. The Physical Attributes category concerns the characteristics of materials as well as visual characteristics of paintings such as artistic color, brushwork and composition techniques. The Styles and Periods category includes commonly accepted terms for stylistic groupings and distinct chronological periods that are relevant to art, architecture, and the decorative arts. The category of Agents includes terms for designations of people, groups of people and organizations involved in possession and selling works of art. The Activities category encompasses areas of physical, mental actions and processes such as archaeology, analyzing and exhibitions. Lastly, the

Materials category includes a variety of materials that could be used in the artwork, while the Objects category contains the concepts referring to various human-made objects used to describe artwork content and the type of artwork itself. Examples of concepts under the Objects category are paintings, amphorae, facades, cathedrals, Brewster chairs, gardens etc. Greenberg [1993] compared several arts-oriented ontologies and found that specific terminology of AAT allows for greater retrieval precision and elimination of unwanted recall.

ULAN (United List of Artist Names) contains information about artists that includes name variants and important biographical information such as dates, locations and historical period. It lists 220,000 artists.

The ontologies discussed above serve as a structural representation of domain-specific knowledge of art domain, where the concepts inter-link and reinforce each other. This representation relates visual, historical, cultural and other types of information. Using ontologies, we can annotate paintings with a large set of concepts, in addition to assigning several well-known terms such as artist name, date and country. In our work, we aim to benefit from the arts ontologies: we utilize artistic concepts and relationships among them to enhance the annotation accuracy of machine learning methods and provide the end users with flexible and meaningful vocabulary of concepts. In the next section, we review the existing user studies of retrieval task in the painting domain. They include discussions of possible strategies for arts images querying, categorizing retrieval concepts and establishing their usability from the point of view of different user groups.

3. 2 User Studies in Paintings Domain

Art is one of the subject fields in which images are used comprehensively, and researchers have extensively analyzed image indexing and retrieval in this field. Brilliant [1988] and Enser et al. [1992] pointed out that many artists and experts in the field use a rough sketch to describe their requirements pictorially. However, Enser et al. [1992] and Garber et al. [1992] recognized that the use of a sketch alone is not sufficient due to the variety of possible interpretations. Garber et al. [1992] pointed out that an art image retrieval system should facilitate explicit descriptions of image contents. Several studies [Panofsky, 1962; Garber et al., 1992] concluded that arts system should ideally facilitate retrieval by a combination of various visual attributes (color, texture), high-level concepts (art period, location) as well as querying by image sketches or layouts.

Several studies focused on the analysis of query concepts for art images. Enser et al., [1992], Jorgensen [1995], Fidel [1997] and Layne [1994] provided a valuable foundation for arts retrieval systems. These classifications include both syntactic (low-level) and semantic (high-level) attributes and differ mostly in the level of detail. Jorgensen [1995] developed the most comprehensive classification of the user queries in the domain of paintings. Table 3.1 shows 12 image classes developed by Jorgensen. Among others, Jorgensen's classification includes visual elements, abstract concepts and art-historical information as useful query concepts in arts domain.

Attribute class	Description
Literal object	Named objects that are visually perceived, e.g., body parts, clothing
People	The presence of a human form
People-related attributes	The nature of the relationship among people, social status, or emotions
Art historical information	Information related to the production context of the image, e.g., artists, medium, style
Color	Specific named colors or terms relating to various aspects of color
Visual elements	Elements such as composition, focal point, motion, shape, texture
Location	Both general and specific locations within the image
Description	Descriptive adjectives, e.g., wooden, elderly, or size, or quantity
Abstract concepts	Attributes such as atmosphere, theme, or symbolic aspects
Content/story	A specific instance being depicted
External relationships	Relationships to attributes within or without the image, e.g., similarity
Viewer response	Personal reaction to the image

Table 3. 1. Jorgensen's classification of image queries

Several studies have focused on the relationships between query concepts and user backgrounds. Hastings [1995], Chen [2001] and Smeulders et al. [2002] grouped users into novice and expert user groups. Smeulders et al. [2002] pointed out the relationship between the user's background and the textual descriptions for the painting provided to him/her. For instance, expert users do not require an explanation of the artifact itself, while a novice user would want to know high-level synopsis about the visual concepts and paintings techniques as well as art historical information such as artist name, painting style etc. Chen [2001] focused on the novice user group and reported the following useful concepts for querying: artist name, historical period and culture, location (indoor/ outdoor), painting style, subject and theme of the paintings. Hastings [1995] performed analysis of the query concepts employed by the expert users. This study found that artist name, abstract concepts, text within paintings (signature) and visual elements (color, brushwork and composition) are useful for the expert user group.

The user studies in arts domain demonstrate that useful query concepts include a wide range of information, including the concepts referring to visual, abstract properties and high-level information. They recognize that the users fall into two broad categories of novice and expert users. Based on these findings, we employ artistic concepts to annotate and retrieve of paintings. In the proposed framework, we recognize the needs of the expert and novice user groups and employ those concepts that have been shown to fulfill the information needs of these groups.

Annotation and retrieval of image contents has largely been addressed in the research community by numerous systems proposed to index and retrieve general domain images. In contrast, annotation and retrieval of artistic images is a relatively new research area. Since artistic images are a subset of general imagery, the existing annotation and retrieval techniques offer one straightforward solution to solve the problem of annotation and retrieval in arts domain. In the next sections we review existing research for efficient indexing and retrieval of general images.

3. 3 Image Retrieval

Since 1970's, image retrieval has been a well-studied topic due to the need of efficient browsing and search through vast image collections. It combines the efforts of two large research communities: information retrieval and computer vision. These communities study the image retrieval task from two different angles. The information retrieval community introduces the text-based paradigm, while the computer vision community focuses on the visual-based paradigm for image retrieval. In this section, we review these paradigms and give some examples of existing image retrieval systems.

3. 3. 1 Text-based Image Retrieval

This very popular framework for image retrieval has two major parts: first, to annotate images with text concepts and then employ the text-based information retrieval techniques to perform image retrieval [Chang et al., 1992]. However, its practical use has two major difficulties that have become more apparent with the growth of the size and versatility of image collections. First, substantial manual effort is needed to prepare the image collections for retrieval. Second, human annotations of images are often inconsistent and imprecise due to the fact that objects within an image simultaneously carry different semantics. For example, an image with tiger can be given such annotations as “tiger”, “animal”, “wild life” and many others. The imprecision in annotation may lead to significant mismatches during the retrieval stage.

3. 3. 2 Content-based Image Retrieval

The two difficulties faced by manual annotation in the text-based approach lead to an alternative approach to image retrieval. Instead of using manually annotated keywords as the basis for retrieval, it was proposed to index image collections based on its visual contents. The typical visual contents include color, texture, structure and shape. This approach established a general framework for content-based image retrieval (CBIR).

Content-based image retrieval systems include three major components: feature extraction, high dimensionality reduction and retrieval design. Feature extraction is concerned with the representation of images within a retrieval system. Generally, features may include both high-level text-based features like keywords and low-level visual features like color, texture and shape. Within the visual feature scope, the features can be further classified into general and domain-specific. The former includes color, texture, while the latter is application-dependent and may include, for example, man-made structures or fingerprint. High dimensionality problems arise from the fact that the number of visual features used can be very high. Since, dimensionality reduction for retrieval systems is not a focus of our research, we refer the reader to the following studies [Minka et al., 1996; Chang and Li, 2003].

The retrieval systems design is concerned with the image querying modes that aim to facilitate effective retrieval in image collections. In their user studies Holt et al., [1995] and Jorgensen et al. [1998] found that the end users experience difficulties while querying the retrieval systems using low-level visual features. These features have limited power for content-based retrieval and their performance is usually application-specific. Since, a typical user does not have the basic knowledge of feature extraction, she is unable to use the system effectively without prior training. The need to express the semantic concepts using adequate features becomes more evident if the image collection includes a large variety of images such as animals, natural scenes, object close-ups, indoor etc. For example, while querying for images with buildings, it is more meaningful to query based on the texture rather than based on color. In contrast, if the user searches for images of plants and greenery, it is more meaningful to query by green colors and texture. Clearly, the retrieval results largely depend on the ability of the user to identify the most expressive subset of features for a query. To make interaction between the user and the system more natural, several querying modes have been proposed. Chang et al. [1998] gave a taxonomy of the existing querying modes; they include:

- Random browsing
- Search by example
- Search by sketch

- Search by text (keywords)
- Navigation using image categories

Despite the variety of retrieval modes offered, user studies [Graber et al., 1992; Holt et al., 1995; Jorgensen et al., 1998] found that search by text is probably the most desirable mode of image search and a combination of several modes like search by text and search by image has the highest usability to the end user. These findings placed importance on the image auto-annotation systems. They led to a current trend in CBIR systems, where image retrieval represents a two-step procedure: first, the user kick-starts the search using semantic concepts and then she interactively looks-up for images [Wang et al., 2001].

In the next section, we focus on the general low-level features used in modern image retrieval and auto-annotation systems. We demonstrate the use of these features in the review of the state-of-art CBIR systems presented in Section 3.5.

3. 4 Image Features

Numerical representation of image content or image features serves as the basis for image retrieval, indexing and annotation tasks. Each image is represented as a feature vector that describes various visual cues such as color, texture and shape within an image database. Given a query image, the system retrieves the most similar images to the query image based on appropriate distance metrics in the feature space. Pavlidis et al. [1978] broadly classified the feature extraction methods into two large groups: spatial information preserving and non-preserving. The spatial information preserving methods derive features that preserve spatial information within an image. Hence, using the extracted features we are able to reconstruct the original image, which makes these methods useful for image compression tasks. Well-known examples of such methods are Principal Component Analysis (PCA) and Independent Component Analysis (ICA). The non-preserving methods aim to represent the image for the purpose of further discrimination. They include color histogram and moments, Tamura texture, Gabor-based texture features, wavelet-based features etc.

Nowadays, almost all annotation and retrieval systems utilize color, texture and shape features for adequate representation of images. The use of multiple image attributes arises from the fact that the use of single image features often leads to a lack of discriminatory power in the annotation and retrieval systems. In this section we briefly review existing methods for extraction of color, texture and shape information in images.

3. 4. 1 Color

Color features are used in a majority of annotation and retrieval systems. Color space and color resolution are important parameters of color extraction methods. Ideally, a color space should be uniform, compact, complete and natural. RGB color space, which is widely used for image representation, does not meet these criteria. Due to this, a majority of annotation and retrieval systems utilize CIE L*u*v color space [Hall, 1988, Chua et al., 1998], which meets these criteria. It is composed of three components, where L defines the luminance and u and v define the chrominance. HSI is another color space that aims to model human color perception, however it is non-linear. Furht [1998] studied the performance of the retrieval system using different color spaces and concluded that while no color space performs best in all cases, the use of color extraction methods in CIE L*u*v and HSI color spaces yields betters retrieval results as compared to that of RGB.

Probably the most popular method for color representation is color histogram. It is generally invariant to translation, rotation and normalized histograms are scale invariant. However, this method is spatially non-preserving. Hsu et al. [1995] observed that visually different images might have similar color histograms. To address this problem, several new representations that account for the spatial distribution of color within an image have been developed [Chua et al., 1998; Vailaya et al., 1998]. Examples include color coherence vector (CCV) [Pass et al., 1996], color region model [Smith et al., 1996] and color pair model [Chua et al., 1994].

3. 4. 2 Texture

Visual texture is defined as a variation of image intensities in the form of repeated patterns [Tuceryan et al., 1993]. These patterns may result from the physical properties of the surface (peakness, roughness) or from the color reflectance. Most images exhibit some form of textures, which provides useful cues for automatic image annotation. In paintings domain the surface of painting provides the cues on the type of brushwork used. Well-known categorization of texture extraction models by Tuceryan et al [1993] includes four major classes. Statistical methods characterize texture in terms of spatial distribution of grey values. This class includes the co-occurrence methods [Jain et al., 1995] and autocorrelation features. Model-based methods assume the underlying model for the description and synthesis of texture patterns. The well-known methods in this class include fractals [Petland et al., 1984] and random field models [Besag 1974]. Geometric methods view texture as being constructed of elements or primitives. Voronoi tessellation features [Tuceryan et al., 1993] and the texture primitives [Blostein et al., 1989] are examples of geometric methods. Signal processing methods utilize the frequency analysis of an image to represent texture. These methods include Fourier domain filtering [Coggins et al., 1985], Gabor filters [Majunath et al., 1996] and Wavelet models [Mallat et al., 1989]. A number of studies [Majunath et al., 1997, Wang et al., 2002] demonstrated that the use of Gabor filters and Wavelet models outperforms the other texture methods in content-based image retrieval and annotation for general image domain.

3. 4. 3 Shape

Shape is one of the most complex visual cues due to the fact that depth information is difficult to acquire from a single viewpoint. Further, object overlapping changes the shape of objects that leads to significant difficulty in object recognition tasks. Various schemes have been proposed for shape representations. These include the string representations [Cortelazzo et al., 2004; Huang et al., 1994], polygons [Schettini 1994], edge direction histograms and moments [Jain et al., 1998] and relaxation techniques [Davis, 1979]. A major disadvantage of the shape representation methods is the fact that a majority of them are not invariant with respect to image size, position and orientation. In order to incorporate rotation and translation invariance, these methods need to cater for all possible positions and orientation, thus increasing the dimensionality of the feature space.

3. 4. 4 Summary of the Low-Level Features

In this section, we summarize the low-level features along with their advantages and limitations. The main objective behind the choice of low-level features for CBIR systems is to ensure appropriate representation of image contents. In terms of color, the most popular features are color histograms [Swain et al., 1991], color moments [Jain and Vailaya, 1995] and color coherence vectors [Pass et al., 1996]. These features describe the global content of image and are easily extracted. Popular shape representations include polygonal approximation [Schettini, 1994], invariant moments [Jain et al., 1998] and Fourier descriptors [Chellappa et al., 1984]. These features require good segmentation algorithms to extract objects from the image. Since objects may be of different scale, orientation and position, the image search using shape features becomes more expensive as compared to search using the color features. In the current CBIR systems, shape features are not used very often because their performance is highly application-dependent. Similarly to shape features, texture features have high complexity of matching.

3. 5 Existing CBIR Systems

In recent years, a large variety of CBIR systems has been proposed. However, systematic studies involving actual users in practical applications need to be done to compare such systems. Here, we discuss the most representative systems and their characteristics.

3. 5. 1 CBIR Systems in General Image Domain

QBIC [Flickner et al., 1995] is the first commercial content-based retrieval system. It supports querying by image examples, user-provided sketches, and color and texture patterns. This system employs mean color and k-element color histogram in RGB, Lab and Munsell color spaces [Faloutsos et al., 1993] to represent color and improved Tamura method [Tamura et al., 1978] for texture. To represent shape, the authors used simple geometrical features.

Photobook [Petland et al., 1996] consists of three image sub-sets, from which shape, texture and face are extracted respectively. The authors employed a 'society of models' approach that accounts for the subjectivity of user perception.

Netra is a prototype image retrieval system developed by Ma and Manjunath [1997a]. The main research contributions of Netra include the use of Gabor filters [Ma and Manjunath, 1996; Manjunath and Ma, 1996], thesaurus construction based on neural networks [Manjunath and Ma, 1997] and image segmentation based on the edge flow method [Ma and Manjunath, 1997a].

MARS (Multimedia Analysis and Retrieval System) was developed at University of Illinois [Mehrotra et al., 1997]. The main focus of MARS is to develop techniques that organize low-level visual features into a meaningful retrieval architecture, which dynamically adapts to different situations. The research contributions include integration of DBMS and IR techniques (exact match with ranked retrieval) [Ortega et al., 1998] and the relevance feedback architecture for query refinements and feature weighting [Rui and Huang, 1998].

SIMPLIcity [Wang, 2000] is a region-based image retrieval system developed at Stanford University. This system introduces and implements semantic image retrieval. This system first classifies the query image into one of the predefined semantic classes such as indoor-outdoor, graph-photograph etc. Next, the system enhances the retrieval results by searching among images under the pre-defined class.

3. 5. 2 Retrieval Systems for Painting Images

Inspired by the growing number of general-domain image retrieval systems. Lewis et al. [2004] proposed an image retrieval system for arts objects. Similar to QBIC, they proposed content-based retrieval using a sample image to query the system. They employed the multiscale color coherence vector to represent color and wavelet-based features using Daubechies filters to represent texture. Recently, they introduced retrieval by extending the functionality of the system with retrieval by crack patterns [Abas et al., 2002]. However, due to the semantic gap between low-level features and human perception, these systems have limited usability since they facilitate image-by-example querying. In our work, we aim to annotate image with actual keywords and, thus, increase usability of the proposed system.

Latest paintings retrieval systems employ domain-specific knowledge to index collections. The significance of these studies is due to the fact that domain-specific knowledge facilitates indexing by a meaningful set of semantic concepts. For example, the retrieval systems developed by Corridoni et al. [1998] and Lay [2004] facilitate querying by semantic color concepts. To index images, these studies employ artistic color theories that define widely known artistic concepts such as *warm* and *cold colors, color harmony* and various types of contrasts using artistic color sphere. Both systems perform back-propagation of region colors onto an artistic color sphere and derive semantic concepts based on it. The proposed systems mostly differ in the image representation and feature extraction methods. Corridoni et al. [1998] performed image segmentation using K-means clustering in CIEL*u*v* color space. To deal with the problem of granularity, the authors represented the image as a multi-level pyramid. In this pyramid, each subsequent level contains image segmentation results based on the iteratively increasing *K*. However, to represent the region colors, the authors utilized mean color. While this approach is adequate for the representation of the Medieval paintings, it is not suitable for the Modern Art, where the authors employed small patches of contrasting colors to give an overall impression. In contrast to this system, Lay et al. [2004] performed the extraction of semantics for each individual pixel followed by the integration of the pixel-based information using expert rules. However, the use of rules imposes scalability concerns. In our work, we employ the Itten's sphere to perform the color analysis and at the same time we avoid the drawbacks of the above-mentioned works.

3. 6 Statistical Learning in Image Domain

These systems employ various techniques to narrow down the semantic gap between low-level features and semantic concepts and enhance the retrieved results. First, through the use of relevance feedback in the image retrieval systems. This technique aims to capture user preferences and provide more accurate results using this information. Second, it is the use of semantic indexing and its close relative, automatic annotation methods. These methods quickly gained research interest since they facilitate concept (or text)-based retrieval in a straightforward manner in contrast to the relevance feedback techniques. Here we review the methods proposed for automatic image annotation.

The major task of image annotation is how to associate the image content (features) with high-level semantic concepts [Chang, 2002]. With the advent of powerful computers, automatic and semi-automatic annotation of image collections using high performance machine learning methods became possible. These methods increasingly employ statistical models to map low-level features onto semantic concepts. Lew et al. [2003] pointed out that the paramount challenge for learning methods remains the bridging of semantic gap. The task of converting easily computed low-level features to the semantic concepts illustrates the semantic gap. This task implies understanding of the semantics behind the concepts and relationships among them.

There exist two major paradigms to tackle the image annotation task. The first paradigm concerns with the use of relevance models for joint modeling of textual and visual data. This paradigm exemplifies probabilistic (except for LSA models) generative models. The second paradigm represents the categorization approach, where individual classifiers focus on annotation of specific semantics.

3. 6. 1 Joint Modeling of Textual and Visual Data

The idea of joint modeling of words and images has been borrowed from the text domain. This paradigm has been extended to the image domain, where the image is described using text vocabulary and feature vocabulary, resulting in finite *image description language* or *blobs*. Both blobs and words are assumed to be generated by hidden variables or *aspects*, which represent a multivariate distribution over blobs and a multinomial distribution over words. Once the joint word-blob probabilities are learnt, the annotation problem is reduced to a likelihood estimation problem relating blobs and words.

Mori et al. [1999] performed one of the early attempts to perform annotation using relevance models. Duygulu et al [2002] and Barnard et al [2003] proposed the hierarchical aspect model to translate a set of image regions into a set of words. Blei et al [2003] employed a Correlation Latent Dirichlet Annotation model, which assumes that the mixture of latent factors follows Dirichlet distribution. Cross-media relevance models [Jeon et al., 2003] represent a closely related approach that borrows from coherent language models. Lavrenko et al. [2003] proposed a continuous relevance model to avoid the problem of cluster granularity. There are several disadvantages of the joint probability modeling approach. First, these models assume that the segmented regions are precise. Second, the number of regions in images is usually unstable, which leads to the difficulty of establishing an adequate number of aspects in such models. Third, to simplify the joint density characterization, the concepts and blobs for an image are often assumed to be mutually independent [Jeon et al., 2003]. Lastly, this approach requires a large dataset of labeled samples to cover broad variations of image samples. This approach is not very useful for the annotation of paintings for several reasons. First, the segmentation techniques often do not represent brushwork adequately: they often combine several brushwork techniques in a single region. Second, the training datasets in the paintings domain are usually limited. They are insufficient for the estimation of joint probability, which may lead to the significant variance error. Due to these disadvantages, we do not utilize the join modeling approach in our work. Instead, we employ categorization approach discussed in the next session.

3. 6. 2 Categorization Approach

The second paradigm is based on categorization. Both generative and discriminative models are used to perform the categorization task in image domain. This approach proposes the extraction of specific semantics: a set of training images with and without the concept of interest is collected and a binary or multi-category classifier is trained to detect the concept of interest. Numerous studies adopted this approach. Examples include detection of people and animals [Forsyth et al., 1997], buildings [Li and Shapiro, 2002], indoor and outdoor scenes [Szummer et al., 1998], cities and landscapes [Vailaya et al., 1998] and trees [Haering et al, 1998]. More recently in paintings domain Herik et al. [2000] and Li et al. [2004] performed annotation of artist names. The learning algorithms used include naïve Bayesian classifier [Keren, 2004], SVM [Feng et al., 2004b] and neural networks [Herik et al. 2000; Breen et al., 2002]. Recent advances in the categorization approach include representation of each concept using mixture models. Thus, multi-category classification model becomes a collection of mixtures [Carneiro et al., 2005; Shi et al., 2006]. These approaches aim to detect an explicit semantics. They require smaller datasets as compared to the joint probability approach. However, the required datasets are still large. Also, these works perform annotation of a flat concept set without account for their internal relationships. In our work, we extend this approach: capture a set of specific keywords by taking into account the relationships among the concepts, while aiming to minimize the number of required training instances using the semi-supervised methods.

3. 6. 3 Semi-supervised Learning Methods

Traditionally both relevance models and image categorization methods follow supervised machine learning framework, where the hypothesis space is constructed based on labeled training samples. However, due to the difficulties of gathering manually labeled data, semi-supervised methods have been proposed. In this section, we review well-known semi-supervised methods used by the state-of-art image annotation systems.

3. 6. 3. 1 Semi-supervised Classification Methods

Using easily available unlabelled data, semi-supervised classification methods modify or reprioritize hypotheses obtained from labeled data alone. The use of unlabelled data leads to higher accuracy of annotation under certain assumptions such as adequate models, features, kernels and similarity functions. Detection of bad matches in advance is a hard problem that remains open [Elworthy, 1994; Cozman et al., 2003]. Semi-supervised learning methods are closely related to the transductive learning paradigm. However, not all semi-supervised methods are truly transductive. In theory, transductive learning methods work on observed data and are not able to handle unseen data. Instead of constructing a general function that handles classification of all instances, these methods extract N observed neighborhood instances and construct a decision function based on these instances for each testing data sample. For example, semi-supervised agglomerative clustering methods are transductive. Notice that under this definition, Transductive SVM and mixture models are in fact inductive, since the classifier is defined over the whole space.

Major semi-supervised learning methods include generative mixture models, self-training and co-training, TSVM and graph-based methods [Seeger et al., 2001]. The generative model approach assumes a mixture of distributions, for example, Gaussians [McLachlan and Basford 1988]. Several authors [Castelli et al., 1995; Castelli et al., 1996; Ratsaby et al., 1995; Cozman et al., 2003] demonstrated that if the model assumption is correct, unlabelled data is guaranteed to improve accuracy of mixture models. This approach has several convenient properties.It represents a class as a number of mixture components. This representation is suitable to represent brushwork techniques, artist names and painting styles since the visual appearance of patterns in each class is non-uniform. For example, in our case we assume that each class of brushwork is represented as a mixture of Gaussians. Nigam et al. [2000] applied mixture models with the Expectation Maximization algorithm [Dempster et al. 1977; Mitchell, 1997] for text classification task. Carson et al. [2002] and Rui et al [2004] employed mixture models for clustering. Baluja [1998] used a similar approach to discriminate face orientations. Debreko et al. [2004] and El-Yaniv et al. [2004] proposed a transductive inference framework based on mixture models for image annotation task. In our work, we further extend the work of these authors to perform feature selection during model construction and utilize available domain knowledge.

Self-training is a commonly used technique, where the classifier iteratively increases its labeled dataset using unlabelled examples that are predicted with high confidence during previous iterations. Several studies applied self-training for text classification [Yarowsky, 1995; Riloff et al., 2003; Maeireizo et al., 2004] and object detection in images [Rosenberg et al., 2005]. However, the problem of this approach is the propagation of error. Due to this, many authors utilize co-training, where a final decision is achieved by combining predictions from two independent sources.

Co-training [Blum and Mitchell, 1998; Mitchell, 1999] assumes that features can be split into two independent sets that are sufficient to train good classifiers that teach each other. Nigam and Ghani [2000] compared co-training with generative models. Goldman et al. [2000], Zhou et al. [2005a; 2005b] and Balcan et al. [2005] proposed different variations of the co-training method. Feng et al. [2004a] proposed a co-training framework with active learning for annotation of large-scale image collections. The disadvantage of this approach is that not every task has two independent sets of features. If the feature sets are not independent then this approach is similar to the self-training approaches. For example, in our work we have only one modality and due to this the application of co-training is problematic.

Several researches focused on transductive SVM methods [Bennett et al., 1999; Fung et al., 1999; Joachims, 1999] that aim to maximize a linear boundary margin on both labeled and unlabeled data. Transductive SVM has been widely used for text classification tasks [Joachims, 1999], however they are not widely used in the image classification.

Graph-based semi-supervised methods define a graph, where the nodes are data points and graph edges reflect similarity among them. These methods are non-parametric and transductive. Well-known examples of graph-based methods include min-cuts method [Blum and Chawla, 2001], harmonic functions for image segmentation [Grady et al., 2004] and Spectral Graph Transducer [Joachims, 2003]. In our future work, we would like to explore these methods.

3. 6. 3. 2 Semi-supervised Clustering Methods

Semi-supervised clustering methods are closely related to the semi-supervised classification. In these methods, labeled data samples serve as *must-links* (two points must be in the same cluster) and *cannot-links* (two points cannot be in the same cluster). There is a tension between satisfying these constraints and optimizing the original clustering criterion, for example, minimizing the squared distances within clusters. Among many methods for clustering, probably the most widely used are *distance-based techniques*. One common characteristic of distance-based clustering techniques is that they assign membership of data points based on the inter- and intra- cluster distance in the feature space. The distance-based clustering approach includes *partitioning relocation* and *hierarchical clustering* techniques [Berkhin, 2002]. Partitioning relocation techniques, for example *K-means,* aim to iteratively relocate data into several subsets. Hierarchical clustering methods iteratively merge (or split) the most appropriate cluster(s) based on the proximity measure called a *linkage metric*. Major inter-cluster linkage metrics [Olson, 1995] include S*ingle-Link*, *Average-Link*, and *Complete-Link*. Recent works that employ semi-supervised clustering include Demiriz et al. [1998], Dara et al., [2002], Bilenko et al. [2005], Shi et al. [2005] and Wagstaff et al. [2001]. For a detailed review of existing works please refer to Grira et al [2004]. In our work we experiment with both distance-based and hierarchical clustering methods. The main disadvantage of these approaches is that various datasets require the use of different distance measures. In our work we successfully solve this problem by introducing the model selection step in our classifier.

3. 7 Ontology-based Image Annotation

One of the disadvantages of traditional supervised and semi-supervised inference methods is the lack of account for hierarchical relationships among semantic concepts [Aslandogan et al., 1997, Hyvönen et al., 2003; Yang et al., 2001]. In an attempt to closely mimic human problem solving strategies, various researches introduced hierarchical machine learning algorithms [Barnard et al., Fan et al., 2005 and 2006]. In the context of hierarchical learning, we recognize *atomic* and *composite* concepts. The details of image, visual properties and objects tend to correspond to atomic concepts, which can be recognized using low-level features. Composite concepts tend to be recognized through juxtaposition of atomic concepts in accordance to the domain-specific knowledge. Hierarchical machine learning algorithms first perform annotation of atomic concepts and then utilize this information to annotate composite concepts.

In general, hierarchical machine learning algorithms fall into two categories: algorithms that learn hierarchy from the training set [Barnard et al., 2001] and algorithms that utilize external hierarchy [Fan et al., 2005 and 2006; Petridis et al., 2006]. The algorithms that learn hierarchy from a training set are useful when we do not have any external knowledge or are not aware about the relationships among concepts. This approach has two drawbacks. First, the intermediate concepts might not be meaningful or a set of images that represents an intermediate concept might be incomplete. Second, these methods usually require a large number of training samples. In contrast, the hierarchical learning methods that utilize external ontology employ independent learners to label images in accord to the concept ontology [Fan et al, 2005 and 2006; Breen et al., 2002]. For example, training a model with respect to 10 unrelated concepts in general requires *10*X* samples, where *X* is the number of samples. Suppose we know that these concepts form an ontology with 6 concepts at the lower level and

4 concepts at the higher level. Then, we need only *6*X* number of training samples to train a model for the annotation of the lower-level labels, and annotate the remaining 4 labels by using ontological relationships. Gruber [1993] defined ontology as the shared understanding of some domains of interest, which is often conceived as a set of classes (concepts), relations, functions, axioms, and instances. Often ontology (or concept hierarchy) is defined as *directed acyclic graph* $G = (N,E)$ that consists of a set of nodes N and set of ordered pairs or *edges* $(N_p,N_c) \in E \subseteq \{NxN\}$. The direction of an edge is defined from the N_p parent node to the child node N_c; this relationship is specified through relational operator $N_p \rightarrow N_c$. Using concept hierarchy, the complete task of concept learning from images is split into several hierarchical subtasks or layers $\{L_1, L_2 \ldots L_n\}$, where each layer is defined as:

$$\{F_l, K_l, T_l, ML_l, h_l\},$$

where l denotes individual layer; F_l denotes the input vector of the relevant feature for layer L_l, K_l denotes the set of concepts relevant to layer L_l., T_l denotes the set of training samples used for the learning subtask, where each element of T_l represents the correspondence between input feature vector and output concept. ML_l denotes a machine learning algorithm that generates a hypothesis h_l, which maps F_l onto K_l based on T_l. The use of pre-defined concept ontology within the annotation system is attractive due to several characteristics. The first is the modularity of concepts. It facilitates the use of several classifiers and features subsets depending on the analyzed concepts. The second is the support of navigation task since the ontological relationships among concepts offer a context for navigation.

3. 7. 1 Existing work

In this section, we review existing studies that focus on the ontology-based annotation for imagery and video. There are several ways to categorize these works. Several studies employ domain ontologies for the concept propagation task. Traditionally these studies assume an initial set of concepts and develop techniques for the annotation using ontological relationships based on the concept propagation [Schreiber et al., 2001 and 2002; Hollink et al., 2003]. However, these works merely use the relationships among the concepts to extend the annotated set of labels. Others focus on the development of ontology for multimedia [Petridis et al., 2006; Saathoff et al., 2006] and information sharing via integration of several ontologies [Soo et al. 2002 and 2003; Dong et al., 2006]. Works of Petridis et al. [2006] and Saathoff et al. [2006] proposed the use of multimedia ontology that serves the needs of learning and retrieval of multimedia information. This approach aims to consolidate visual attributes with the general and domain-specific ontologies. Current implementations combine the existing metadata standards [DCMI, 2001; McBride et al., 2004; Manjunath et al., 2002] with domain-specific ontologies. Studies of Soo et al. [2002 and 2003] integrate domain-specific ontologies with the RDF standard [McBride et al., 2004] to make collections easily accessible. These works are more useful for the annotation of general images since paintings domain includes several well defined domain ontologies such as AAT, ULAN and ICONCLASS. These ontologies can be further extended and combined with the general purpose ontologies such as the WordNet. Numerous works focused on concept disambiguation using ontologies [Fan et al., 2005a and 2005b; Bilenko et al., 2005; Srikanth et al., 2005; Shi et al., 2006]. Often these studies employ domain-specific ontologies to introduce meta-level information [Fan et al., 2005a and 2005b]. Similar to this work, we utilize meta-level information to annotate high-level semantics. To perform annotation and disambiguation, the proposed methods often employ distance-based clustering techniques [Bilenko et al., 2005; Srikanth et al., 2005; Petridis et al., 2006] and probabilistic methods [Shi et al., 2006; Fan et al., 2005 and 2006]. In our work, we experimented with both above-mentioned methods to perform annotation and further extended them to facilitate disambiguation based on the ontological relationships among concepts. Existing works include both automatic and semi-automatic efforts in ontology-based annotation. These are semi-automatic annotation using ontology-based annotation tools [Schreiber et al., 2001 and 2002; Hollink et al., 2003], automatic approaches for concept propagation [Breen et al., 2002] and concept disambiguation [Fan et al, 2005; Srikanth et al., 2005]. In our work, we aim to develop an automated framework for concepts annotation and disambiguation.

Traditionally, the concepts of ontology are represented using text. However, in multimedia context it makes sense to include visual examples to “teach” a system regarding the membership of unlabelled samples. The majority of studies that perform concept disambiguation employ multi-modal ontologies, since they associate a subset of training data with the concepts of ontology. Similar to this works, our ontology is multi-modal: annotated concept is associated with some visual examples.

3. 7. 2 Advantages of Hierarchical Concept Representation

When looking at an image, we can understand it and easily identify atomic and composite concepts that can be used for its description. The studies discussed above demonstrate that this task can be easier for machine learning if we introduce hierarchical concept organization within the inference process. In our work, we perform automatic annotation of paintings based on domain ontology, where the visual-level information serves as meta-level for the annotation of high-level semantics and ontological relationships serve to disambiguate automatically generated labels. Our work has the following advantages that arise from the use of domain ontology:

- Guide for manual annotation

Annotation template based on the hierarchical structure ensures consistent manual annotation of the collection, thus reducing potential ambiguities due to the annotators prior knowledge.

- Good extensibility

  Since atomic concepts serve as the basis for the annotation of large number of high-level artistic concepts, the hierarchical structure incorporates the newly added high-level concepts without the need to rearrange already existing ontology concepts.

- Bridging the gap between atomic and composite concepts

  Since hierarchical concept structure includes relationships among concepts, it is possible to induce high-level composite concepts through atomic concepts. Figure 3.1 demonstrates the example. Here, assuming that the system correctly identifies that the painting exhibits *mezzapasta* and *shading* brushwork classes, *primary* color palette and *chiaroscuro* contrast, it can then deduce *Medieval* period of art using hierarchical concept relationships.

Figure 3. 1. Girl with a Pearl Earring, by Johannes Vermeer

- Account for the concept relationships

  There exist several strategies depending on the type of the relationships. The concept relationships can be used to minimize the number of the required classifiers for such cases as synonymous concepts. Next, the concept relationships facilitate concept disambiguation. Further, concept relationships offer rich context for navigation in contrast to the traditional keyword-based approach, which suffers from the so called “too many or nothing” problem [Chang et al., 1998]

- Easily extended to audiovisual media

  Many ontology-based annotation methods implicitly introduce visual information into domain ontology since they relate a concept with a set of training samples. In general, we can derive concepts more accurately with the help of multi-source information as compared to single-source information.

3. 8 Existing Problems and Research Directions

Despite the progress made in various aspects of image annotation and retrieval, there are still numerous research issues that need to be solved to successfully implement retrieval systems for arts images. In our work we aim to propose a framework that would minimize these concerns:

1. Minimizing the need for labeled dataset.
2. The use of domain knowledge for annotation.
3. The handling of user heterogeneity.
4. The use of additional information sources.

3. 8. 1 Minimizing the Need for Labeled Dataset

The majority of machine learning techniques require consistent manually annotated set of training samples. In paintings annotation task, there exists a large number of images, making manual annotation task erroneous, time consuming and costly. In many cases, each painting is assigned multiple labels representing its visual, factual and abstract content. Further, each painting requires two independent sets of labels: for its blocks as well as for the whole image. This makes annotation process very tedious and requires extensive expertise of the human annotator.

There are several research paradigms that address this problem. The first arises from the fundamental property of the statistical inference methods: by minimizing the number of numerical features used for the inference, we minimize the required labeled dataset. Due to this property, feature selection methods become important for the machine learning task. The second paradigm is concerned with the use of unlabeled samples during the inference process. Semi-supervised learning addresses the following questions: 1) Can we combine a relatively small labeled set and a large unlabeled set and achieve the same accuracy as the fully labeled set? 2) Can we increase the accuracy of annotation by using a combination of labeled and unlabelled instances as compared to using only labeled instances? In this dissertation, we are primarily interested in these two questions that facilitate reduction of the required training dataset while retaining reasonable accuracy of learning. We propose the novel transductive inference framework that performs feature and model selection and demonstrate that the use of this framework ensures higher annotation accuracy.

3. 8. 2 The Use of Domain Knowledge for Annotation

As we have discussed in Section 3.7.3, the use of concept ontologies in image annotation systems offers several benefits. Recent trends include the use of general domain ontologies or domain-specific ontologies for the annotation task. The use of domain ontology is beneficial, since it facilitates concept disambiguation and propagation as well as more natural navigation and retrieval. Our work is different from the existing painting annotation works [Herik et al., 1998; Li et al., 2004] since it utilizes domain knowledge of paintings domain to support the auto-annotation task. The major question is how to incorporate the structural domain knowledge within the inference framework. In this thesis we propose an annotation framework to: 1) perform both region-based and image-based annotation of paintings; 2) incorporate the ontology concepts and their relationships to induce high-level semantic concepts; 3) perform robust classification of concepts at various levels of granularity; and 4) incorporate domain knowledge to disambiguate artistic concepts. There are other important research issues related to the use of domain ontologies. The first is the development of the retrieval systems that facilitate ontology-based query construction and navigation. The second is the use of RDF and other standards to relate the domain ontology to the existing arts-oriented ontologies and publish the annotated collections online. We plan to focus on these directions in our future work.

3. 8. 3 Handling User Heterogeneity

A relatively small body of annotation systems research recognizes that the information needs of the users are not similar due to variations of user's background. The ability to account for user backgrounds is especially desirable in specific imagery domains such as artistic, geographical and medical imagery. In our work we utilize domain ontology that caters to the informational needs of a wide range of users. We recognize the expert and novice user groups in paintings domain, since these groups possess different knowledge about artistic concepts. Ideally, the annotation framework should account for the needs of various user groups.

3. 8. 4 The Use of Additional Information Sources

In the experimental setting, we often assume a fully automated system without human input. However the real life is quite different, since the system can obtain cues about high-level semantic concepts from user actions [Jain, 1993]. One popular method is relevance feedback (RF) [Smith and Chang, 1997; Rui et al., 1998]. Alternatively, it is possible to utilize the World Wide Web to extract necessary information. In terms of human interaction, it is very prolific and, thus a more promising source for extraction of semantic content. There exist at least two strategies to utilize WWW. First, we can extract textual annotations that accompany millions of images posted online. The analysis of the free text posted by the users has the potential to solve the problem of manual labeling. Second, it is possible to engage the users to perform online annotations. Recent trend in this area is the use of social networks, where the users are invited to feel themselves as experts and perform annotation of images within game-like scenarios [von Ahn et al., 2004]. In this thesis, we perform preliminary experiments with partial annotations. We plan to focus on the use of partial annotations in more detail in our future work.

Chapter 4

Ontology of Artistic Concepts in the Paintings Domain

4. 1 Introduction

Traditionally, artistic concepts serve as one of the major tools for the description, categorization and navigation in the domain of western painting collections [Arnheim, 1954; Canaday, 1981; Itten, 1961; Lazzari, 1990; Pumphrey, 1994]. Artistic concepts vary vastly in their scope. They include concepts referring to the detailed pictorial information such as impasto brushwork class, various abstract characteristics such as expressive, gestural and concepts used for retrieval applications such as painting style, artist name etc. Artistic concepts represent a wide range of high-level concepts for paintings retrieval that describe paintings in various levels of detail. Table 4.1 demonstrates examples of queries with artistic concepts.

The artistic concepts, which characterize pictorial information, represent the visual language of paintings. The artists employ the visual language to describe style of artists and paintings styles, periods in fine art and various abstract characteristics. For example, *complimentary* palette, *impasto*, *divisionism* or *scumbling* brushwork classes with *complimentary* contrasts represent the *post-impressionism* paintings style.

Accounting for a wide range of artistic concepts is beneficial to painting retrieval for several reasons. First, it facilitates flexible retrieval of arts images at various levels of granularity. The end user is able to query the retrieval system with high-level and specific artistic concepts. For example, queries such as "*Medieval* paintings with *shading* brushwork class in *cold temperature*" become possible. Second, annotation of specific artistic concepts facilitates spatial retrieval of paintings. The system facilitates retrieval based on the queries like "Paintings with *scumbling* brushwork class on top and *chiaroscuro* contrast". Third, it offers a novel application of query by example paradigm. This is important for queries that are easily expressed by visual means. The user is able to submit an image and query "Paintings with similar distribution of color temperature". Lastly, it offers a basis for automatic comparison of paintings. For example, the system is able to decide that painting A has stronger complimentary contrast than painting B. Thus, the user is able to submit a query image and retrieve images using queries like "Paintings with stronger chiaroscuro contrast".

1. Paintings with *impasto* brushwork class in red color;
2. Paintings with *complimentary* palette and *temperature* contrast;
3. Paintings with *scumbling* brushwork class on top and *chiaroscuro* contrast;
4. *Expressive* painting with *mezzapasta* brushwork class and *complimentary* color contrast;
5. All paintings with *optical mixing*;
6. Paintings with *wet on dry* and *warm temperature* colors;
7. Paintings in *impressionist* painting style;
8. *Medieval* paintings with *shading* brushwork class in *cold temperature*;
9. Paintings by *van Gogh*;
10. *van Gogh's* paintings in *warm* colors;
11. Paintings by *Cezanne* with *impasto* brushwork class and *temperature* contrast;
12. *Modern* art *expressive* paintings

Table 4. 1. Examples of queries based on artistic concepts

In our framework, we organize artistic concepts of various levels of detail within a three-level ontology. This ontology is meant to support and facilitate flexible annotation and retrieval of paintings. In the next section, we discuss this hierarchical concept structure and its levels in detail.

4. 2 Three-level Ontology of Artistic Concepts

To perform annotation, we organize artistic concepts into an ontology that combines concepts into three inter-linked levels: visual level, abstract level and application-specific level. Figure 4.1 depicts the three-level ontology of artistic concepts.

Visual level of the taxonomy includes concepts that refer to the visual properties of paintings such as color and brushwork, composition, materials, type of medium and other classes. In our current work, we focus on the color and brushwork classes, while we aim to incorporate the other visual attributes within the annotation framework in our future work. The visual-level concepts are beneficial for two reasons. First, these concepts serve as common basis for the retrieval of paintings. As discussed in Section 4.2, retrieval by visual-level concepts is useful for the expert user group. Second, visual-level concepts facilitate disambiguation of high-level concepts due to the fact that they are related to various high-level artistic concepts.

We employ visual-level concepts as meta-level information for mapping from low-level features to high-level concepts. In Sections 4.3.1 and 4.3.2 we will discuss these concepts in detail.

The next level of taxonomy includes abstract concepts. These concepts refer to non-visual information available in paintings. They include perceptual properties and general terms referring to brushwork and colors. These serve primarily the expert users for navigation and retrieval purposes. In Section 4.4 we will focus on the abstract-level concepts.

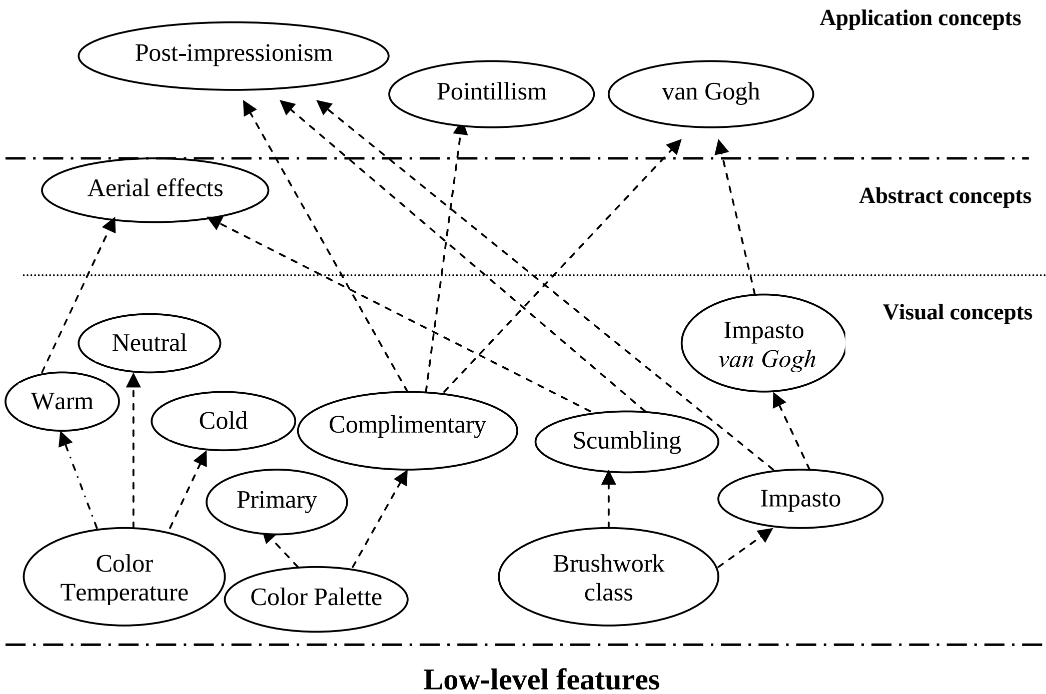


Figure 4. 1. Three-level ontology of the artistic concepts

The third level of taxonomy contains application-level concepts. This level includes high-level concepts such as artist names, painting styles and art periods. These concepts are used for navigation and retrieval by the novice users and widely offered in virtual galleries and museum websites. Similarly to the abstract-level concepts, the visual-level concepts are related to the application-level concepts and serve as cues for their definition based on the domain knowledge. We will discuss the application-level concepts and demonstrate examples of their definitions based on visual-level concepts in Section 4.5.

The three-level ontology of artistic concepts includes relationships between the concepts ar various levels as well as relationships within each level. This ontology combines the concepts from AAT, ICONCLASS and ULAN ontologies since each of them offers a different view of the visual information. These ontologies borrow definition of artistic concepts from various art historical studies such as works of Arnheim [1954], Itten [1961], Canaday [1981], Lazzari [1990], Pumphrey [1994] and many others. Since AAT, ULAN and ICONCLASS ontologies are meant for manual annotation, they do not explicitly define relationships between concepts of visual level and concepts of abstract and application level. These relationships are implicitly defined in the free-text definitions of high-level concepts. Due to this, these ontologies readily facilitate mapping of low-level features onto visual-level concepts and mapping of low-level features onto high-level concepts without accounting for visual-level concepts. However, direct mapping of low-level features onto high-level semantic concepts does not always result in satisfactory performance and raises scalability concerns for large paintings collections as demonstrated in the experiments of Li et al. [2004]. In contrast to these ontologies, the three-level ontology of artistic concepts facilitates a bottom up approach, where visual concepts serve as intermediate steps for learning application-level concepts. Such organization of artistic concepts mimics domain knowledge for auto-annotation of images to a higher extent as compared for AAT, ICONCLASS and ULAN ontologies.

Overall, explicit representation of concepts in visual, abstract and application-level concepts offers more flexible retrieval, rich context for navigation, facilitates comparison of paintings and links to the widely-known art ontologies AAT, ULAN and ICONCLASS. In the rest of this section, we focus on individual levels of the three-level concept ontology.

4. 3 Visual-level Artistic Concepts

Artists utilize visual language for paintings description and categorization [Arnheim, 1954; Canaday, 1981; Itten, 1961; Lazzari, 1990; Pumphrey, 1994]. Table 4.2 demonstrates the list of visual-level concepts we employ in our work.

In the Western paintings domain, the major visual language concepts characterize color, brushwork and composition. This list is by no means exhaustive, but it represents widely used concepts that provide cues for the annotation of high-level concepts used by novice and expert user groups.

VISUAL-LEVEL CONCEPTS	DESCRIPTION
Color	
Color Palette	Specific set of colors used by artists. Three major concepts include primary, complimentary and tertiary palette.
Color Temperature	Perceptual property of colors. Green-blue-purple hues define *cold* color temperature, orange-yellow-red define *warm*; violet and yellow-green hues define *neutral* temperature.
Contrast	Three types of color contrasts widely used by artists. Complimentary contrast measures the contrast between color hues. Various artistic theories arrange color hues in circular order such that the directly opposite hues represent the strongest contrast. Temperature Contrast denotes the contrast between colors of different temperature. "Warm-cold" pair represents the strongest temperature contrast. Chiaroscuro Contrast is the contrast between two colors in terms of their intensity (shading).
Brush	
Brushwork classes	Brushwork classes denote various techniques of brush application. In the three-level ontology we include brushwork classes widely used in western paintings. In terms of the surface, the artists distinguish from washed flat techniques to thick opaque techniques. In terms of color, a brushwork patch exhibits single to multiple color hues.

Table 4. 2. Artistic concepts of the visual level

4. 3. 1 Color Concepts

Analysis of artistic color concepts constitutes a large body of expert analysis in the paintings domain. These concepts represent intermediate information that facilitates the annotation of painting styles, art periods, and to some extent, artists. For example, *light-dark* contrast is used in the *Renaissance* period, while the artists working in the *Post-impressionism* and *Fauvism* painting styles used *complimentary* contrast in their works [Canaday, 1981]. Similarly, certain color expressions characterize artists. For instance, Berezhnoy et al. [2004] analyzed the usage of *complimentary* contrasts by *Vincent van Gogh*. This type of contrast uniquely characterizes this painter. Due to this, the authors employed color contrast analysis to automatically establish authenticity of van Gogh's paintings.

We investigate artistic color concepts based on the theory formulated by Itten in 1961 [Itten, 1961]. This theory proposes the mapping between colors and artistic color concepts, and is primarily used by artists. Itten defines twelve fundamental hues and arranges them in *color circle*. Color circle is an artistry color model. Unlike RGB, CMY, and HSI, which are used primarily to facilitate color specification, a color circle transcends the constructive objective of color specification to also represent artistry color relationships. It is a specifically tuned color space whose geometrical arrangement exhibits relationships articulated in the theory of color contrast and harmony.

Figure 4.2 demonstrates the arrangement of colors on the artistic sphere. A color circle consists of three *primary* colors, three *complimentary* colors and six *tertiary* fundamental hues. Fundamental hues of color circle vary through five levels of intensity and three levels of saturation, i.e. 15 levels for each color. Each fundamental hue serves as the basis for such variations, thus creating a subset of colors.

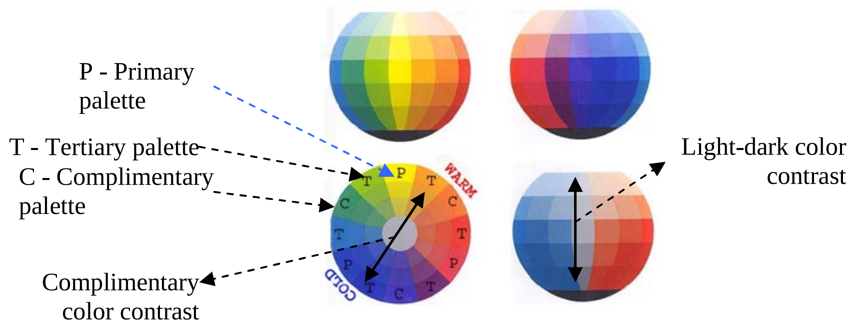


Figure 4. 2. Itten's chromatic sphere

The total set of colors derived from the color circle contains 180 colors that are organized as a chromatic sphere. Fundamental colors are arranged along the equatorial circle of sphere, luminance varies along medians and saturation increases as the radius grows. Itten located the shades of gray colors in the center of the sphere and white and black colors at the poles of the sphere.

Colors of the artistic sphere with yellow-red-purple fundamental hues have *warm* color temperature, while colors based on green-blue-violet hues have *cold* color temperature. *Neutral* color temperature characterizes colors based on green-yellow and red-violet hues. These colors may change their *neutral* temperature to *cold* or *warm* depending on the surrounding colors. Figure 4.3 demonstrates paintings in warm and cold colors.

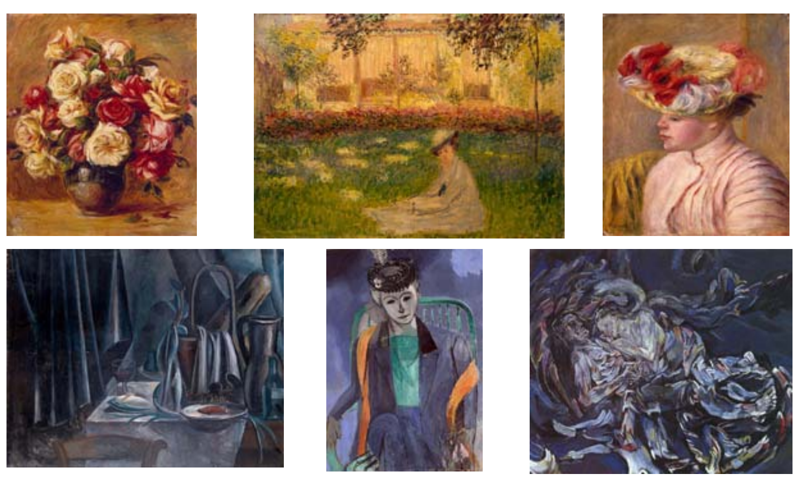


Figure 4. 3. Examples of color temperature concepts.
Paintings in warm and cold colors are in upper row and lower row respectively

The artists categorize color palette into *primary*, *complimentary* and *tertiary*. Primary palette represents the set of yellow, red and blue fundamental hues; complimentary palette represents the set of colors with green, orange and violet fundamental hues; and tertiary palette represents the other six fundamental hues of the color circle. As all colors of the chromatic sphere except black, white and grays, are derived from the fundamental hues, each color exhibits the same color palette category as its respective fundamental hue. Properties of colors such as intensity and saturation influence the perceptual appearance of color temperature and color palette. Color temperature and color palette are most apparent in the fundamental hues of the color circle, and its appearance gradually decreases with the changes in intensity and saturation towards the poles of the sphere.

Color contrast is a relative measure defined for at least two colors. Following Itten, we perform analysis of the four well-known color contrast types: *complimentary*, *light-dark*, *temperature* and *value* contrasts. *Complimentary* contrast represents relationships between fundamental hues. Figure 4.4 demonstrates examples of complimentary contrast in paintings.

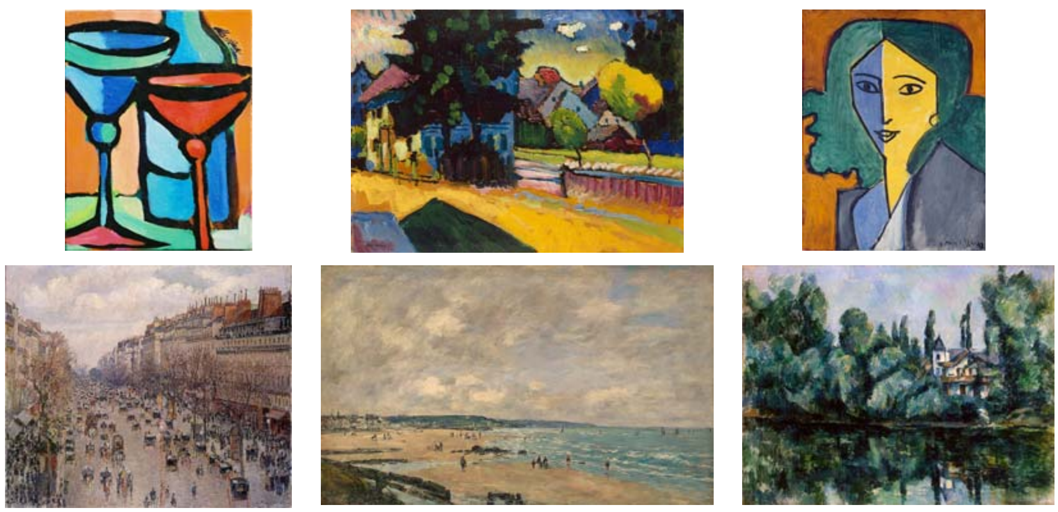


Figure 4. 4. Examples of complimentary contrast

Paintings with high and low degree of complimentary contrast in the upper and lower rows respectively

Directly opposite hues of color circle have the strongest complimentary contrast. *Light-dark* color contrast accounts for the difference in color intensity of two colors, while value contrast reflects the difference in color saturation. *Temperature* contrast reflects the interaction of different color temperature patches. Itten defined interaction of warm and cold color temperatures to be stronger color temperature contrast as compared to the interaction of neutral and warm or neutral and cold color temperatures. *Value* contrast reflects the difference of saturation between two colors.

4. 3. 2 Brushwork Concepts

Brushwork refers to the pattern depiction with the help of brush [Pumphrey, 1996]. Brushwork has various properties such as length, width, jitter, opacity etc. In the field of art, the combination of such brushwork properties defines the *brushwork technique* or *brushwork class*. Together with the color concepts, experts employ brushwork classes to analyze and describe the paintings [Canaday et al., 1981]. Various studies exploited the fact that brushwork provides strong cues to the painting style and artist name. For example, in the area of image synthesis, Hertzmann [1998] manipulated length, jitter and opacity of brushwork to synthesize images in various painting styles such as *Impressionism*, *Expressionism* and *Post-impressionism* painting styles. Various researches [Li et al, 2004; Herik et al., 2000] employed brushwork to perform classification of paintings with respect to artist names.

There are two major approaches to automatic brushwork analysis. The first approach focuses on the explicit detection of brush-strokes and assessment of their properties. The works of Kropatsch et al. [1995], Meltzer et al. [1998] and Sablatnig et al. [1998] serve as examples of this approach. In these works, the authors developed methods for the detection of single and overlapping brush-strokes for further identification of artists. This approach has several drawbacks. First, it makes a number of assumptions regarding the brush-stroke intensity, size and shape. Second, is requires a controlled high-resolution collection.

The second approach performs indirect assessment of brushwork properties via texture-based representation and analysis of brushwork patches. Works of Herik et al. [2000] and Li et al. [2004] exemplify this approach. It has significantly lower computational complexity as compared to the explicit detection of brush-strokes. Further, explicit detection of brush-strokes is problematic due to brush-stroke overlapping; especially in painting styles of *Modern* art period. Lastly, texture-based analysis of brushwork is expected to perform better for non-controlled collection. For example, for the collections downloaded from the Web like in our case. Thus we focus on the texture-based analysis of brushwork in this thesis.

We perform analysis of brushwork using eight widely known brushwork classes that dominated in western paintings from the 10^{th} up to 19^{th} century [Lazzari, 1990; Canaday, 1981]: *divisionism*, *glazing*, *grattage*, *impasto*, *mezzapasta*, *scumbling*, *shading* and *pointillism*. Table 4.3 demonstrates examples of these widely known brushwork classes with their short description and prominent characteristics. *Divisionism* denotes the application of regular small touches of unmixed contrasting colors so that they combine optically. This brushwork class represents the *color mixing principle* widely used in the *Modern* period of art.

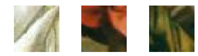
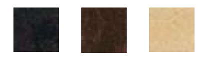
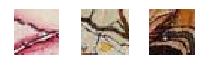
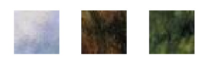
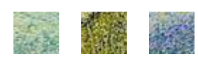
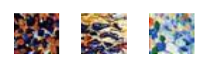
Class	Background	Characteristics	Examples
Shading	Depiction of foldings in Medieval Period	Edges and gradients, often directional, intensity contrast, weakly or non-homogeneous	
Glazing	Depiction of nudity/face in the Medieval Period	Subset of hues (yellow, red, orange), intensity contrast, gradients, non-homogeneous, may contain edges	
Mezzapasta	Widely used technique in paintings. The color palette used varies with respect to the art period.	Homogeneous, low intensity contrast and small gradients	
Grattage	Depiction of objects and patterns in Fauvism and Expressionism painting styles of the Modern Art	Edges, high gradients, intensity contrast, inhomogeneous	
Scumbling	Depiction of sky, clouds, greenery in Fauvism, Impressionism, Post-impressionism and Pointillism painting styles of the Modern Art	Soft gradients, low intensity and hue contrast, low directionality, weakly homogeneous	
Impasto	Widely used in Impressionism, Post-impressionism, Pointillism	Edges, high gradients, often directional, low hue contrast, high intensity contrast	
Pointillism	Often used for depiction of atmosphere/air in Pointillism painting style	Medium intensity contrast, medium roughness, no directionality, homogeneous	
Divisionism	Widely used in Pointillism, demonstrates the Color Mixing Principle	High gradients, high roughness, high intensity and hue contrast, no directionality, weakly homogeneous	

Table 4. 3. Examples of brushwork classes

Glazing has been used mostly in the *Medieval* and less in *Modern* periods. This technique represents a thin layer of transparent paint to highlight soft gradients and inner glow. It is primarily used for portraits and nudity depiction.

Grattage brushwork class was invented in the *Modern* art period and found mostly in paintings of *Fauvism* and *Expressionism* painting style. The brushwork class denotes the use of sharp lines to depict an object.

Impasto has been widely used in a variety of painting styles and periods, but mostly in *Post-Impressionism* painting style of the *Modern* period of art. Paintings by *Vincent van Gogh* exemplify this technique to the highest extent. This brushwork class represents the use of opaque and thick layers of paint with characteristic ridges due to the sliding of a brush.

Mezzapasta class is widely used in the *Medieval* and *Modern* periods. It represents brushwork patches with plain smooth color. This brushwork class is often used to color large areas or backgrounds in painting.

Scumbling is used in various styles of the *Modern* art, but mostly in the paintings of *Impressionism* painting style. It represents a series of unorganized overlapping strokes in different directions to create objects like clouds, hair, water and grass.

Shading is mostly used in the *Medieval* period of art for the depiction of folds on clothing. It represents directed series of flat long strokes of uniform color.

To perform the annotation task, we represent brushwork as a set of mutually exclusive classes. Thus, each pattern in our dataset belongs to only one class of brushwork. However, several properties of brushwork significantly complicate the annotation process. First, brushwork patches might bear some resemblance to each other. For example, *divisionism* in some paintings is similar to *impasto* brushwork class. Second, brushwork varies significantly in the areas along object borders and areas of minor details. Further, our collection includes paintings captured under varying lighting conditions and this introduces additional difficulty. Third, each brushwork class includes a variety of patterns since it includes patterns of this brushwork class by various artists, from various painting styles and art periods. Figure 4.5 explains this phenomenon using the diagram of *impasto* brushwork class as an example.

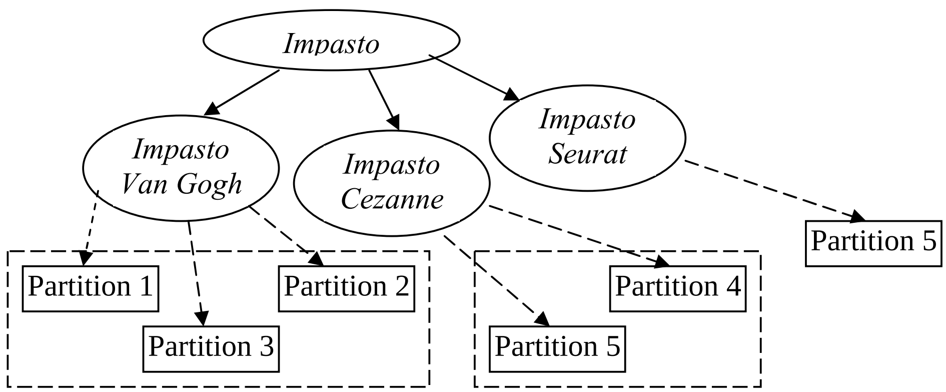


Figure 4. 5. An example of pattern distribution in the *impasto* brushwork class

The diagram includes three levels. First level contains clusters of patterns in the feature space (partitions). Second level represents clusters of brushwork class patterns with respect to artists. For example, *impasto by van Gogh* has more bold ridges and opaque colors, while *impasto by Cezanne* has more fine frequent ridges and relatively more transparent colors etc. Lastly, the third level combines all these specific representations into a more general brushwork class *impasto*. Further the diagram underlines the importance of accurate brushwork class detection. Figure 4.5 demonstrates that brushwork provides cues to accurately predict artist name via impasto by van Gogh, impasto by Seurat clusters. Overall, brushwork meta-level concepts compliment the meta-level color concepts, which provide limited cues for the prediction of artist names, and together they serve as intermediate information for the auto-annotation of high-level concepts such as artist name, painting style and period of art.

4. 4 Abstract-level Artistic Concepts

This level includes high-level concepts that are widely used by experts for painting description and retrieval. Often these concepts represent perceptual characteristics of paintings due to the specific use of colors and brush [Itten, 1961; Lazarri, 1990; Canaday, 1981]. Table 4.4 represents the list of abstract concepts used in the three-level ontology. The artist names in parenthesis next to concepts denote which artists are known to use the respective concept often.

- **Warm (All artists)**
- **Cold (All artists)**
- **Expressive**
    - **Complimentary Color Contrast (Seurat, Matisse)**
    - **Temperature Color Contrast (Van Gogh,Cezanne)**
- **Chiaroscuro**
    - **Light-dark Color Contrast (Rubens, Vermeer)**
- **Rational**
    - **Divisionism (Seurat, Cezanne)**
    - **Pointillism (Matisse, Pisarro)**
    - **Shading (Vermeer, Caravaggio)**
    - **Glazing (Rembrandt, Rubens)**
    - **Mezzapasta (Matisse)**
- **Gestural**
    - **Impasto (Van Gogh, Cezanne)**
    - **Grattage (Matisse)**
    - **Scumbling (Cezanne, Monet, Pisarro)**
- **Aerial Effects**
    - **Scumbling (Cezanne, Monet, Pisarro)**
    - **Pointillism (Matisse, Pisarro)**
- **Optical mixing**
    - **Scumbling (Cezanne, Monet, Pisarro)**
    - **Divisionism (Seurat, Cezanne)**
- **Wet On Dry**
    - **Scumbling (Cezanne, Monet, Pisarro)**
    - **Divisionism (Seurat, Cezanne)**
    - **Pointillism (Matisse, Pisarro)**
    - **Glazing (Rembrandt, Rubens)**
    - **Shading (Vermeer, Caravaggio)**
    - **Grattage(Matisse)**
- **Wet On Wet**
    - **Mezzapasta (Matisse, Delacroix)**
    - **Impasto (Van Gogh, Cezanne)**

Table 4. 4. Heuristics definitions for the abstract-level concepts

The definitions of *warm* and *cold* abstract concepts are borrowed from Itten's color theory [Itten, 1961]. They refer to the use of *warm* and *cold* color temperature throughout whole canvas.

Chiaroscuro [Itten, 1961] represents *light-dark* contrast in the canvas with the use of predominantly skin-like and brown hues. This concept is widely used by *Leonardo Da Vinci*, *Rembrandt* and *Rubens*.

Arnheim [1954] defined *rational* and *gestural* concepts. *Rational* concept includes brushwork classes that require the careful application of brush-strokes such as *divisionism*, *pointillism* etc., while *gestural* groups brushwork classes where brush-strokes are applied in uncontrolled gestures.

The concept *expressive* has several definitions in terms of the use of color, brush and content itself. We employ the definition by Itten [1961], who defined *expressive* as the use of *complimentary* or *temperature* color concepts in the canvas.

Aerial effects and *optical mixing* are central to the *Modern* period of art [Lazarri, 1990]. *Aerial effects* include brushwork classes that aim to achieve sensation of air in paintings. Such brushwork classes are *scumbling* and *pointillism*. *Optical mixing* refers to the placing of contrasting colors next to each other such that from the distance it creates new color.

The artists classify all brushwork techniques in the domain of western paintings by the method of application: *Wet on Dry* or *Wet on Wet* [Canaday, 1981]. *Wet on wet* denotes the blending of colors together while the first application of paint is still wet. The artists mostly use these techniques to create the areas of homogeneous color. *Wet on Dry* concept refers to the application of color over the dry coat of color underneath.

From Table 4.4 we observe, that visual-level color and brushwork concepts are related to a large number of abstract-level concepts. Such relationships represent one of the benefits of the ontology-based annotation discussed in Section 3.7. They offer convenient basis to perform annotation of the abstract-level concepts without training additional classifiers. Having the visual-level concepts assigned, it is possible to exploit the concept relationships and perform concept propagation to annotate abstract-level concepts.

Next, application-level concepts (in this table, artist name) are indirectly related to the abstract-level concepts via visual-level artistic concepts. Due to this, we do not employ these relationships for the annotation task. Overall, the application-level concepts are useful for flexible querying and navigation in the ontology-based system.

4. 5 Application-level Artistic Concepts

In this section we discuss the last level of the three-level ontology. It includes high-level concepts used by novice users such as artist name, period of art, painting style, movement and country. In our study we focus on painting styles, historical periods and artist names. Such concepts as movement and culture can be extracted from the ontology using their relationships with painting style, historical period and artist name concepts. For the annotation task, two types of relationship between artistic concepts are of intersect. First, ontological relationships between visual-level and application-level concepts. Second, relationships among application-level concepts.

Similar to the abstract-level concepts, domain-specific knowledge [Canaday, 1981; Pumphrey, 1996] defines the application-level concepts using rule-based heuristics based on the visual-level concepts. Table 4.5 demonstrates examples of such heuristics for Impressionism, Fauvism and Pointillism painting styles.

VISUAL-LEVEL CONCEPTS	**APPLICATION-LEVEL CONCEPTS**		
	IMPRESSIONISM	**POST-IMPRESSIONISM**	**POINTILLISM**
Color			
Cold temperature	Nan	Nan	Nan
Warm temperature	Nan	Nan	Nan
Neutral temperature	Nan	Nan	Nan
Primary palette	High	Nan	Nan
Complimentary palette	Nan	High	High
Color Contrast			
- complimentary	Low	High	Low
- temperature	Low	Medium	Low
-light-dark	Medium	Low	Low
Brush			
Impasto	Low	High	Low
Shading	Low	Low	Nan
Grattage	Nan	Nan	Nan
Pointillism	Nan	Nan	High
Mezzapasta	Medium	Low	Nan
Glazing	Medium	Nan	Nan
Scumbling	High	Medium	Low
Divisionism	Nan	Low	High

Table 4. 5. Examples of heuristics for definitions of application-level concepts

Artist name and art period concepts are defined similarly. In this Table, we place visual-level concepts in the leftmost column and employ *High-Medium-Low-Nan* scale to describe it. Nan value denotes that the visual-level concept is not related to the definition of application-level concept. Following domain-specific knowledge, we employ color and brushwork classes to represent heuristics that defines painting style concepts. From Table 4.5 we make several conclusions. First, it demonstrates that visual-level concepts serve as visual cues to distinguish application-level concepts. Second, only a subset of visual-level concepts contributes to the heuristics definition of a particular application-level concept. Lastly, application-level concepts usually exhibit a mixture of the visual-level concepts. For example, the painting styles from Table 4.5 are defined as a mixture of brushwork classes.

We demonstrate relationships between visual-level and application-level concepts using examples in Figures 4.6 and 4.7. Figure 4.6 shows paintings from several art periods and painting styles.

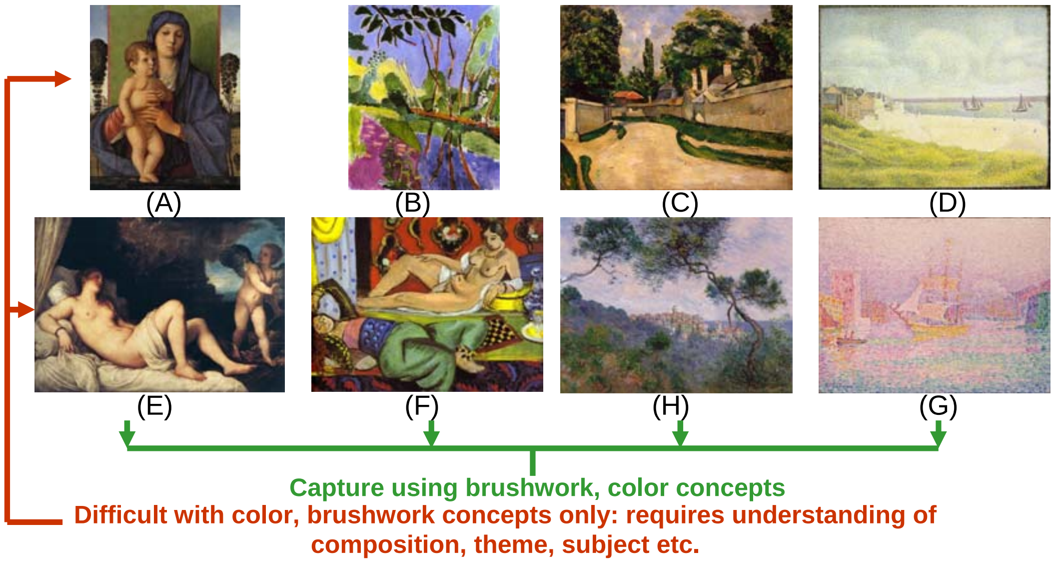


Figure 4. 6. Examples of Painting Styles and Art Periods

In columns from left to right: Medieval Art (Baroque (image A) and High Renaissance (image E) and painting styles of Modern Art (Fauvism (images B and F), Impressionism (images C and H) and Pointillism (images D and G))

It is clear that artistic color and brushwork concepts serve as cues for the recognition of such styles and periods as Pointillism (images D and G), Fauvism (images B and F), Impressionism (images C and H), Modern Art (images B-D and F-G) and Medieval Art (images A and E). However, to recognize *Baroque* and *High Renaissance* (paintings A an E in Figure 4.6) we require additional cues such as composition, theme and subject information. Figure 4.7 demonstrates paintings of artists *van Gogh*, *Cezanne* and *Martina*. Similar to the example above, it shows that visual-level concepts contribute to the recognition of various artists. However, the degree to which this meta-level information facilitates successful recognition varies. For example, it is relatively easy to recognize painting by Martina from paintings by van Gogh and Cezanne, while it is more difficult to recognize paintings of van Gogh from paintings of Cezanne.

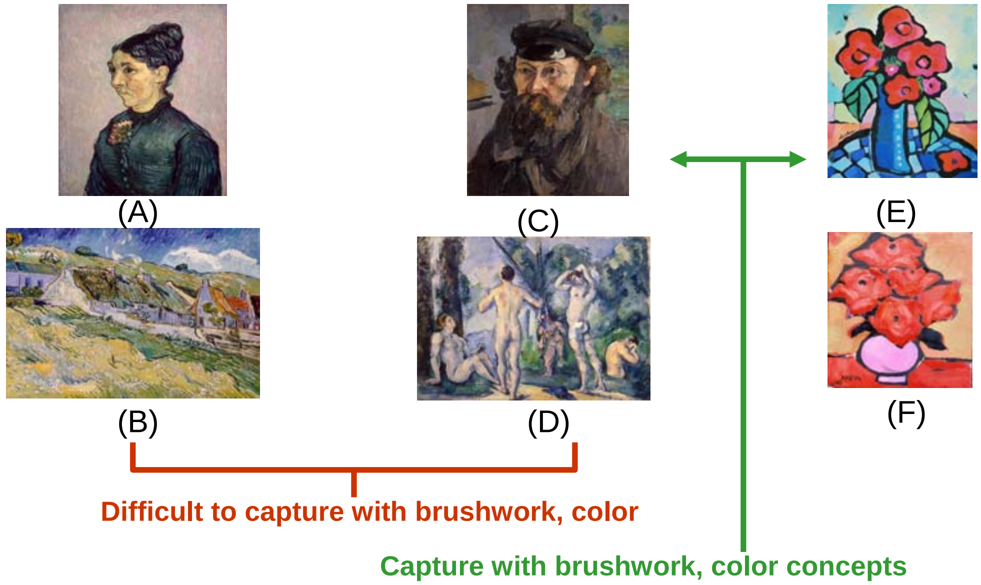


Figure 4. 7. Examples of Artists

In columns from left to right images by van Gogh (image A-B), Cezanne (image C-D) and Martina (image E-F)

Next, we discuss relationships among the application-level concepts. Table 4.6 provides examples of relationships among application-level concepts based on the fine arts timeline from 1250 to 1900 [Canaday, 1981]. It includes the concepts of artist names, painting styles, art periods and era. From Table 4.6 we can see relationships exist between artist and painting style, artist and period, and painting style and art period concepts. As we discussed in Section 3.7.3, these relationships facilitate concepts expansion and disambiguation. In our study, we exploit these relationships for concept disambiguation to ensure that the final labels of artist, painting style and art period are consistent with domain-specific knowledge.

Era	Period	Artist	Painting Style
1250	Medieval	Giotto, Lorenzetti	Gothic
1400		Botticelli, da Vinci, Piero, Lippi	Early Renaissance
1500		Raphael, Titian, El Greco, Bruegel	High Renaissance Northern Renaissance
1600		Rubens, Rembrandt, Poussin, Leyster	Baroque
1700		Boucher, Watteau, Hogarth	Rococo
1750		Fragonard, David	Neoclassicism
1800		Goya, Ingres, Constable	Romanticism, Realism
1850		Bierstadt	Pre-Raphaelites
1875	Modern	Cassatt, Gauguin, van Gogh, Monet, Morisot, Seurat	Impressionism, Post-impressionism, Pointillism, Expressionism
1900		Matisse, Picasso, Dalí, Lange	Abstraction, Fauvism, Cubism, Futurism, Dada and Surrealism

Table 4. 6. Timeline of the western fine art from 1250 to 1900

4. 6 Summary

In this chapter, we introduced artistic concepts that are widely used for the manual annotation of paintings. These concepts represent both visual and high-level semantic information. We organize artistic concepts into a three-level ontology, where visual-level concepts describe pictorial properties of paintings and application and abstract levels include semantic concepts. In accordance to domain-specific knowledge, visual-level concepts serve as cues for the description and annotation of high-level semantic concepts. To employ these cues for automatic annotation, we represent visual-level concepts as meta-level information that facilitates the inference of high-level concepts. In the three-level ontology, we represent two levels of semantic concepts. These are abstract and application levels that support the retrieval by expert and novice user groups respectively.

Overall, the three-level ontology of artistic concepts serves various purposes. First, it describes paintings at various levels of details, thus offering a basis for painting annotation with both high-level and detailed visual concepts. Second, it facilitates concept disambiguation, flexible retrieval and navigation based on the concept relationships. In our work, we employ the three-level ontology of artistic concepts to narrow down the semantic gap for automatic annotation of paintings. In the next chapter, we discuss the proposed framework for ontology-based annotation of paintings with artistic concepts.

Chapter 5

Framework for Ontology-based Annotation of Paintings with Artistic Concepts

5. 1 Introduction and Motivation

Due to the digitization of museum collections, automatic annotation and retrieval of paintings became of practical and research interest [Lewis et al., 2004]. Early works [Flickner et al., 1995] proposed the use of low-level features (visual similarity) to perform retrieval of arts images using query-by-example (QBE) strategy that does not facilitate retrieval based on semantic concepts. Various studies [Holt et al., 1995; Wang et al., 2001] found that the query-by-keyword strategy (QBK) is more useful to the end users. This strategy allows the users to search for images by specifying their own query using a limited vocabulary of semantic terms. Numerous research works [Jeon et al., 2003; Barnard et al., 2003; Lie et al., 2004] proposed semantic indexing of images collections using statistical machine learning techniques. These studies address various aspects of automatic image annotation such as the learning of specific semantics, the use of hierarchical learning methods, adaptive selection of models and many others. However, image annotation task remains challenging due to the fundamental problem of semantic gap and concept ambiguity.

In Chapter 4 we discussed the concept ontologies for paintings domain that are widely used for categorization and navigation of paintings collection. In accordance to the domain-specific knowledge, these concepts are organized into hierarchical structure, where more specific visual-level concepts serve as cues for annotation of high-level abstract and application-specific concepts. For example, the use of *scumbling* brushwork class with *complimentary* palette points out that a painting is likely to be of *impressionism* painting style.

In our framework, we aim to utilize such domain-specific knowledge and demonstrate that its use within the annotation framework enhances the quality of annotation. There are several important questions that we need to address in order to tackle the problem of automatic paintings annotation using artistic concepts. First, how to adequately represent color and brushwork information in paintings? Second, what concepts can we learn from images directly and how to organize semantic concepts? Lastly, how to incorporate domain-specific knowledge into the annotation process for the purpose of concept disambiguation and expansion?

An adequate representation of color and brushwork requires to account for several criteria. First, the size and resolution of images influences the representation of color and brushwork concepts. This is especially crucial for brushwork analysis, since ultimately it relies on the intensity distribution within image blocks. A second criterion is the choice of features. The low-level features used should have high discriminative power and facilitate translation of pixel distribution into color and brushwork concepts. These features should account for several important properties of color and texture such as coarseness, directionality, major hues and brightness as well as capture spatial distribution of pixels within a block.

For the second question, the choice of concepts and their taxonomy relies on the domain-specific knowledge discussed in Chapter 4. But not all concepts can be learnt and acquired from an image based on its visual contents using image processing and machine learning techniques. For example, this task will be difficult for such abstract-level concepts as *expressive*, *harmony* etc. In analogy to the general image domain, we recognize atomic and composite artistic concepts. The meta-level visual concepts that encode brushwork and color concepts are atomic. These concepts have relatively consistent visual representation and can be acquired using machine learning techniques. The concepts of abstract and application levels of the three-level concept ontology are composite concepts. They are often represented and perceived as combinations of atomic concepts. Usually these concepts have a wide variety of visual representations. This is the major reason why learning and acquiring these concepts based on low-level features have limited success. To remedy this situation, we aim to utilize atomic concepts and their relationships to composite concepts to perform the annotation task. To mimic human reasoning within the annotation framework, we exploit the three-level ontology of artistic concepts that encodes relationships between atomic and composite concepts. Such organization is natural due to its consistency with cognition rules of human learning, thus resulting in a useful ontology structure. In Section 3.7.3 we discussed other benefits of the hierarchical concept organization.

For the third question, domain-specific knowledge is naturally depicted in the three-level ontology of artistic concepts in Chapter 4. We utilize this ontology in several ways. First, we view the visual-level color and brushwork concepts as meta-level information within the proposed framework. Combined with the low-level features, they facilitate more accurate annotation of various composite artistic concepts. Second, during the annotation process, we employ domain-specific information about similarity of artistic concepts. For example, similarity information about brushwork classes facilitates iterative disambiguation and recognition of classes. Lastly, we utilize ontological relationships to perform disambiguation of high-level composite concepts. Accounting for these relationships helps to reprioritize the system's judgments about candidate concepts and enhance the quality of annotation.

5. 2 Overview of Framework for Ontology-based Paintings Annotation

Annotation of images with high-level concepts is a complex task. To perform annotation, our annotation framework includes three major stages: a) segmenting images into meaningful units of analysis; b) extracting appropriate features for the units and c) mapping these units onto atomic and composite concepts. The problem of annotation can be expressed as:

$$C(I_i) \approx C\,(\,S\,(I_i)\,) \approx C(F_c(R_{ij}) \cup F_b(R_{ij})) \approx C(\cup K(F_a(R_{ij})) \rightarrow L \qquad (5.\ 1)$$

$$S^r(I_i) \approx R_{ij}$$

where $i = \{1 \ldots N\}$ and N denotes the number of samples in a training set. $_j$ denotes region within image $I_{i.}$ L denote the set of concepts. Function $S(I_i)$ refers to a transformation of the content of an image. In our framework, we perform segmentation of image contents into regions/blocks R_{ij}, thus, $S(I_i) \approx \cup R_{ij}$. The function $F_x(R_{ij})$ performs annotation of image blocks R_{ij}, where $F_c(R_{ij})$, $F_b(R_{ij})$ and $F_a(R_{ij})$ perform annotation of visual-level color and brushwork concepts, and application-level concepts respectively. Function K then performs disambiguation of block-level labels, finally, function $C(I_i)$ generates annotation of an image I. As expressed in Equation 5.1, we divide the image contents into units R_{ij}, thus, assuming that the function $C(I_i)$ can be approximated by the union of block-based annotations generated by the functions F.

To facilitate annotation, we aim to utilize the three-level ontology of artistic concepts. This ontology offers atomic and composite concepts for annotation, thus, $L = L_c \cup L_a$, where L_a refers to the set of atomic concepts and L_c refers to the composite concepts. During the annotation process, we exploit relationships between atomic and composite concepts, which are encoded in the three-level concept ontology. First, we perform mapping of low-level features onto concept set L_a using functions $F_c(R_{ij})$ and $F_b(R_{ij})$. Each of the functions returns posterior probability generated by a learner. Then, we combine relevant low-level features and atomic concepts to generate annotations of high-level concepts L_c using function $F_a(R_{ij})$.

Lastly, function *K* performs disambiguation of the generated annotations for image units to achieve annotations for the whole image. While statistical learning is one of the often-used techniques to integrate block-based information and disambiguate concepts [Feng et al., 2004], we utilize ontological relationships to complete this task.

Probably the closest work to ours is the work of Fan et al. [2005, 2006]. In this work, the authors introduce manually constructed domain ontology that includes both atomic and composite concepts.They perform probabilistic inference of atomic concepts, followed by the inference of composite concepts using the conditional probability distribution of atomic concepts. Other similar approaches include the works of Srikanth et al. [2005] and Petridis et al. [2006].Our work is different from their contribution in several ways. First, we propose statistic inference that utilizes domain knowledge at several levels in addition to the domain-specific ontology. Second, in our framework we focus on adaptive selection of features and model parameters as well as minimization of the training datasets required. Overall, our framework includes three major stages: image segmentation and low-level feature extraction, annotation of image blocks/segments with visual meta-level artistic concepts, and annotation of high-level concepts. Figure 5.1 illustrates the proposed framework.

First, we perform image segmentation and represent image regions/blocks using low-level features. Several studies employed block-based or region-based approach to the paintings analysis [Herik et al., 2000; Li et al., 2004]. We experiment with two types of image regions: a) color/texture blobs generated using image segmentation techniques; and 2) fixed-sized blocks (32x32 pixels). In our framework, we represent visual content by image regions/blocks using color, texture and geometrical low-level features. To perform annotation of visual-level color concepts, we employ the artistic color theory of Itten [1961], which offers a mapping between color hues and visual-level color concepts. We demonstrate that by using visual-level concepts and their ontological relationships the proposed method facilitates the annotation of abstract artistic color concepts without additional training. Specifically, we employ the artistic color sphere and fully supervised probabilistic SVM classifier. For effective annotation of brushwork patterns, we adopt the serial multi-expert approach, where sequentially arranged experts (learners) perform step-wise disambiguation of the target concepts based on a decision hierarchy. The decision hierarchy encodes relationships among classes, thus iteratively splitting a dataset into sub-classes until the leaf nodes that model the target concepts are reached. Due to its modularity, this approach facilitates feature selection and model selection for each node of the decision tree. We combine this approach with semi-supervised learning methods to address the problem of limited labeled datasets. Using this method, we investigate: a) one-step annotation of brushwork classes and step-wise disambiguation using multiple experts; and b) manual and automatic selection of low-level features and parameters of the semi-supervised learning methods and the use of distance-based and probabilistic semi-supervised learning methods. We aim to demonstrate that the resulting transductive inference using multiple experts is effective for the annotation of complex brushwork patterns and that the proposed methods for automatic feature and parameter selection technique is comparable to the manually assigned features.

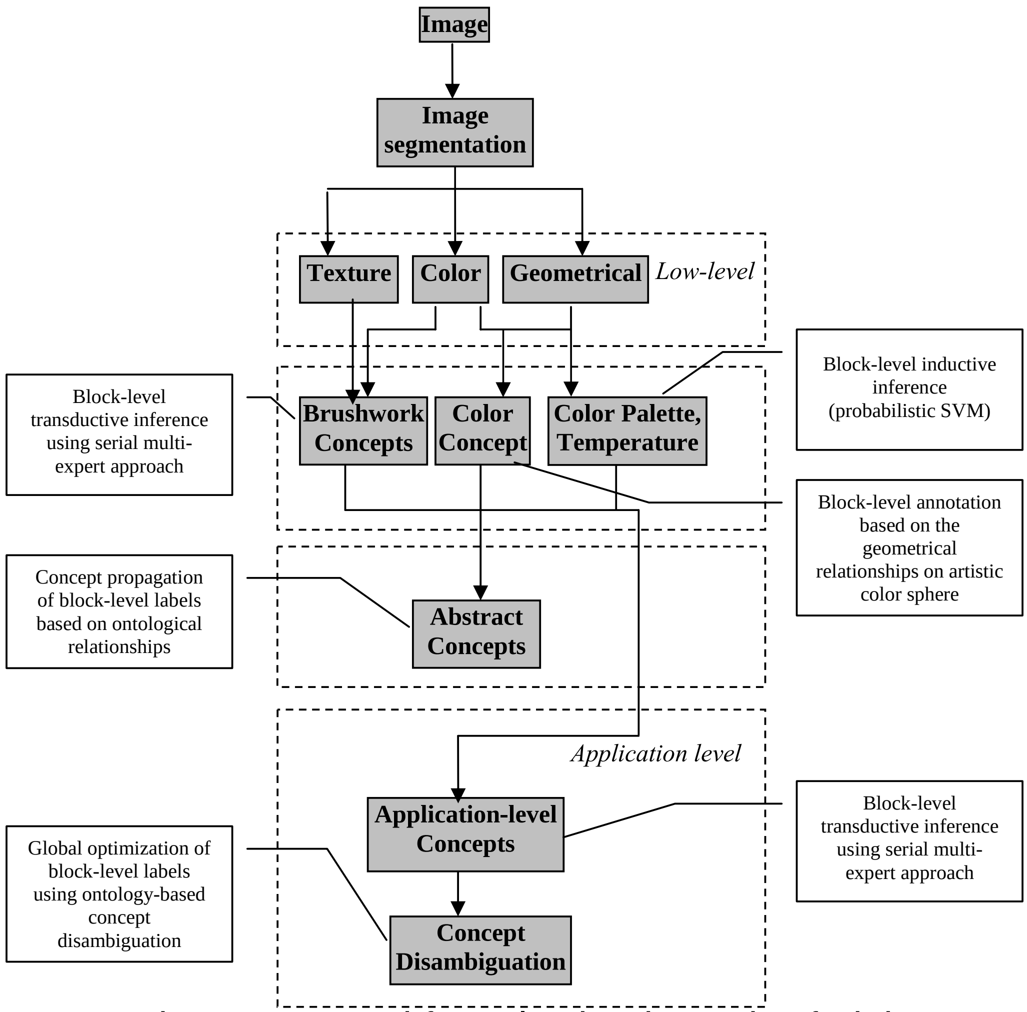


Figure 5. 1. Framework for ontology-based annotation of paintings

Second, we perform annotation of high-level concepts. We distinguish two groups of high-level concepts: abstract-level and application-level concepts. Abstract-level concepts include various semantic terms used by the art experts such as “gestural”, “rational”, “expressive”, “warm” and others. Application-level concepts are meant for the novice users. In our framework, we include artist name, painting style and art period concepts in this level. The distinction between abstract and application levels is due to several reasons: a) these levels facilitate paintings retrieval for different user groups; and b) we employ different approaches to perform annotation of these concepts. To annotate abstract-level concepts, we perform concept propagation based on visual-level concepts due to the fact that heuristic rules for these concepts are clearly defined. The application-level concepts are not defined in such a straightforward manner. To annotate application-level concepts we employ a two-step procedure: a) the annotation of image regions with high-level semantic concepts; and b) the integration of the generated concepts to annotate the whole image. For step (a) we employ the semi-supervised techniques developed for brushwork annotation. In this step we exploit the fact that visual-level concepts serve as cues for annotation of high-level concepts. We thus utilize the visual-level concepts as meta-level information and employ the transductive inference and multiple experts to label the whole image with high-level artistic concepts such as the artist name, painting style and art period. We aim to demonstrate: a) the importance of meta-level information in the annotation process; b) the effectiveness of multiple experts approach as compared to one-step inference approach; and c) the effectiveness of the proposed method to generate satisfactory performance using limited training set. Third, using the generated labels, we further exploit the ontological relationships among high-level concepts to disambiguate concepts.

In this thesis, we mostly focus on the annotation of application-level concepts, since it is easy to test as the ground truth is easily available from the World Wide Web. To explore of abstract-level concepts, we perform several experiments using expert-provided ground truth and aim to focus on these concepts in more detail in our future work.

5. 3 Dataset for the Evaluation of the Proposed Framework

In this section we discuss the dataset we employ for the evaluation of the proposed framework. Table 5.1 shows the details of the dataset collection. It is composed of western fine art paintings in two periods of art, seven painting styles and eleven artists.

This collection includes the most representative painting styles in each period of art and the widely known painters in each painting style. The painters under the same painting style are difficult to distinguish, since they share the similar set of painting techniques. Overall, the dataset includes 1050 paintings. For our further experiments, we split the dataset into two independent subsets: 315 paintings for training purposes and 735 paintings for the testing purposes. In the Table 5.2 we demonstrate the examples of paintings in the dataset.

Period	Painting Style	Artist	Number of Images
Modern	Fauvism	Matisse	84
	Impressionism	Monet	146
		Renoir	138
	Post-Impressionism	Van Gogh	76
		Cezanne	116
	Expressionism	Schiele	150
	Pointillism	Cross	81
		Seurat	78
Medieval	Renaissance	Titian	60
	Baroque	Rembrandt	59
		Frans Hals	62
TOTAL			1050

Table 5. 1. The dataset used for the framework evaluation

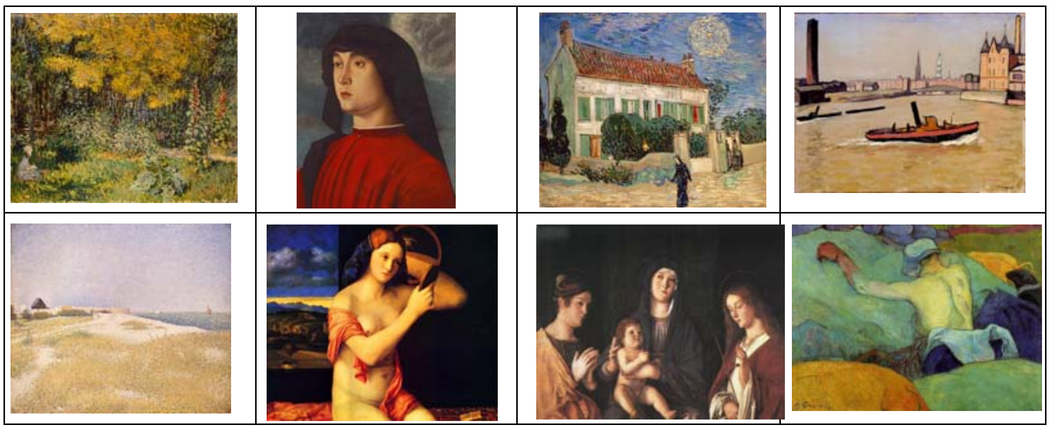


Table 5. 2. Examples of the paintings in the dataset

We compared our dataset with the collections used in the existing works for the annotation of art images. Table 5.3 summarizes these datasets in terms of: reference to the work, the size of the dataset, analyzed categories and the results achieved. In this Table we combine studies that focus on both Western and Chinese paintings domain. This is due to the fact that in all the discussed work the set of features, the nature of the task itself and classification approaches are comparable. Based on this Table, we observe that our collection is larger as compared to the datasets used in the existing studies. It has a large number of categories and comprises different art periods. The right-most column of the Table 5.3 demonstrates the performance levels achieved by existing works. It is clear that the number of the analyzed categories is crucial for both painting style and artist name annotation. The small number of categories and small focused dataset usually result in relatively high annotation accuracy. Also, the performance depends on the categories themselves. Wang et al. [2006] and Li et al. [2003] observed this phenomenon in their works. These authors demonstrated that the experiment results can vary up to 2 times depending on the number of the categories used and their visual similarity. For example, due to highly dissimilar categories, Icoglu et al. [2004] achieved high performance in their recognition. In our work, we observe the same phenomenon: images of artists from the same period and painting style are confused more often when compared to the images from different periods and painting styles.

Work	Dataset Size	Number of categories	Categories	Accuracy
[Jiang et al., 2004]	800	2	Painting Style (landscape vs. flowers)	94%
[Wang et al., 2006]	600	11	Painting Style + Medium	35% to 65%
[Wang et al., 2006]	360	5	Painting Style + Medium	42 to 74%
[Icoglu et al., 2004]	154	3	Painting Style (Impressionism, Cubism, Abstractionism)	90%
[Li et al., 2003]	276	5	Artist (Chinese Art)	62-87%
[Herik et al., 2000]	60	6	Artist (Western Modern Art)	85%

Table 5. 3. Comparison of the dataset with that used in the existing works

5. 4 Summary

In this chapter, we presented our motivation and proposed a framework for automatic painting annotation using artistic concepts. This framework attempts to index paintings with a large variety of artistic concepts for the purpose of flexible querying, retrieval and navigation by end users of different backgrounds. To perform the annotation, the proposed framework relies on domain knowledge: it utilizes the domain-specific ontology during annotation of both visual and high-level artistic concepts. By using domain-specific concept ontology we aim to narrow down the semantic gap between low-level features and artistic concepts. This concept structure is opened and can be augmented with new concepts without sacrificing the system's robustness.

Next, we discussed the dataset for the evaluation of the effectiveness of the proposed framework. While this dataset is small when compared to the general image benchmarks, it is more challenging as compared to the datasets used for the existing works that annotate arts images.

In our work, we aim to develop a fully automatic framework that employs machine learning techniques to annotate images with artistic concepts. We aim to demonstrate that the use of domain-specific ontology has several advantages: 1) the use of meta-level information facilitates higher accuracy of semantic concept annotation as compared to the direct mapping of low-level features onto these concepts; and 2) ontological relationships facilitate disambiguation of the automatically generated annotation and further increase in the system performance.

In the next three chapters, we will discuss different parts of our framework. For annotation of visual-level color concepts, we employ the traditional supervised learning scheme (in Chapter

6). To annotate brushwork patterns, we employ the combination of semi-supervised learning and multi-expert approaches (in Chapter 7). Finally, to demonstrate annotation of high-level concepts and concept disambiguation, we employ the combination of the proposed serial multi-expert scheme and linear programming techniques (in Chapter 8).

Chapter 6

Inductive Inference of Artistic Color Concepts for Annotation and Retrieval in the Paintings Domain

6. 1. Introduction and Motivation

In this chapter we focus on the annotation of image with artistic color concepts that capture a large body of expert knowledge in paintings domain. Annotation and analysis of artistic color concepts has two benefits. First, these concepts serve as meta-level information for annotation and retrieval of paintings with high-level concepts of artists, painting styles and art periods. In the domain of western paintings, combinations of color concepts are known to characterize the artists and painting styles [Berezhnoy et al., 2004].

Second, automatic annotation of color concepts such as color temperature, color palette and color contrast facilitates automated annotation and retrieval of paintings by these concepts in large-scale artwork databases. Recently several systems have been proposed for retrieval in arts databases by such cues as color and texture based on the QBE paradigm [Flickner et al., 1995; Lewis et al., 2004]. Such querying paradigm introduces ambiguity at the query stage. In our work, we propose to index images by the semantic color concepts and facilitate QBK querying paradigm for paintings retrieval.

6. 2 Related Work

The majority of methods for the analysis of color concepts in arts domain utilize artistic color theory such as Itten's color theory [1961] and Munsell color space [Munsell, 1915]. Morphological and geometrical relationships among colors on the artistic color sphere define various artistic color concepts, including color temperature, color palette and color contrasts.

The general pipeline of such methods [Corridoni et al., 1998; Lay et al., 2004; Stanchev et al., 2003] is as follows: split the image into color regions, back-project the mean region color onto the artistic color sphere and utilize relationships among artistic colors to index an image with associated color concepts. Corridoni et al. [1998] and Stanchev et al. [2003] employ the K-means clustering method to split an image into regions. Further, they back-project the mean region color onto the quantized space of artistic colors. However, color averaging leads to a loss of information about the distribution of colors within a region. Such information is desirable for the analysis of *Modern* art paintings styles (*Post-impressionism, Impressionism* and *Pointillism*) and various artists (*Van Gogh, Cezanne, Monet*). In the paintings of these artists, contrasting colors and colors of different color temperature are placed close to each other at the very fine level. Thus, the distribution of artistic color concepts, which pertain to each pixel within the color region, is non-uniform. Consequently, the use of averaged color to assign artistic color concepts does not model the artistic color concepts of a region accurately. Further, the works of Corridoni et al. [1998] and Stanchev et al. [2003] do not account for mutual interaction of various color temperatures. The approach of Lay et al [2004] is somewhat different. The authors performed back-projection based on individual image pixels. To integrate the color temperature, color palette and contrast information, the authors employed a rule-based approach that encodes domain knowledge. The major drawback of this system is the fact that rule-based inference lacks robustness and the knowledge base grows large due to the need to account for various color distributions.

To alleviate some of the above problems, we propose: 1) a representation of image regions with multiple colors; 2) a combination of generic and domain-specific features for annotation and 3) the use of machine learning techniques to mimic human perception of color temperature and color palettes. To facilitate adequate and efficient image retrieval, we perform annotation of image color/texture region. However, several authors [Wang et al., 2006; Li et al., 2003] observed that the color/texture segmentation of images often tends to merge areas of different brushwork within a single region. This results in non-adequate representation of brushwork within a region. Due to this, the authors utilized small fixed-size blocks to perform annotation with respect to high-level semantics. In our work, we employ both the segmented color/texture regions and image blocks. To facilitate efficient representation and retrieval of images by the color information, we employ the segmented image regions. To perform annotation of images with high-level semantics, we need to perform annotation of color and brushwork meta-level information. To perform this task we rely on the fixed-size blocks. In the rest of this chapter, we discuss the proposed method, perform evaluations and summarize our findings.

6. 3 Framework for Annotation with Artistic Color Concepts

We perform automated analysis of artistic color concepts in three steps: image segmentation, analysis of the color concepts at the visual level that characterize image regions, and the analysis of the abstract-level color concepts that characterize the whole image. During this three-step annotation process, we extensively employ domain-specific knowledge, namely Itten's color theory and its major element, the artistic color sphere. We utilize the artistic color sphere for the annotation of image regions. To annotate image regions with artistic concepts, we employ two types of inference: machine learning to annotate color temperature and color palette concepts; and geometrical relationships among artistic colors on the sphere to infer color contrast. We employ supervised learning, since it facilitates account for various properties of a region, including color distribution, which are not discussed in Itten's color theory. However, the use of supervised learning for annotation of color contrast is a difficult task, since it requires data samples for each combination of color hues, brightness and saturation. To perform annotation of color contrast concepts, we exploit the arrangement of colors on the artistic color sphere. Geometrical relationships among colors define the degree of *complimentary, temperature, light-dark* and *value* contrast among them. Later in this section, we focus on our method in detail.

6. 3. 1 Image Segmentation

The analysis of color temperature and contrast concepts requires taking into account the spatial distribution of colors within a painting. Due to this requirement, global representations of color such as color histograms are inadequate for this type of analysis. To facilitate adequate retrieval by color information, we need to account for its size, position and length of the border. To generate such regions we employed a color/texture image segmentation technique. We tested several segmentation techniques such as Blobworld [Carson et al., 2002], Mean-shift [Comaniciu et al., 1999] and the method of Rui et al. [2004]. The Blobworld segmentation method produces the most acceptable results for the analysis of artistic color concepts. This method extracts color/texture features and groups them together using a combination of the Expectation Maximization and Minimum Description Length methods. Similar to the other two segmentation methods, Blobworld does not produce ideal regions but it is relatively more tolerant to the brushwork variance.

6. 3. 2 Color Region Representation

Next, we extract low-level color and geometrical features for each region. Currently each region maintains multiple dominant colors in our system unlike the methods proposed by

Corridoni et al. [1998] and Lay et al. [2004]. We calculate the vector of dominant colors using both CIE L*u*v and HSI color spaces. For each color space, the system calculates a color histogram and normalizes it by its maximum value. Using the top *k* values in the color histogram, we select the *k* dominant colors. To perform projection of dominant colors onto the artistic color sphere, we operate in the CIE L*u*v color space due to its linearity. We convert dominant colors in the CIE L*u*v color space to the corresponding reference colors in the artistic color sphere as follows:

$$ref = \arg_i \min_{1 \le i \le N} dist(R_c, M_c(i)) \qquad (6.\ 1)$$

where *dist* denotes the normalized Euclidean distance, *Rc* denotes the CIE L*u*v* values of a dominant color, *Mc(i)* denotes the CIE L*u*v color values of color i of the artistic color sphere, and N denotes the number of such reference colors ($N = 187$, including 5 shades of gray and black and white colors).

We calculate the geometrical features to facilitate spatial retrieval by color concepts. For this task we account for the region area and its position. In addition, we perform simple morphological operations by extracting contacting border between neighboring regions to store their location and normalized length.

6. 3. 3 Color Temperature and Color Palette Annotation

In this task we are concerned with the distribution of *warm, cold* and *neutral* temperatures within a region, since color temperatures influence each other and their spatial distribution produces a variety of perceptual effects [Itten, 1961]. Figure 6.1 demonstrates the distribution of the color temperature within a block.

Figure 6. 1. Distribution of the color temperature within a block

From left to right here: original block, pixels of cold color temperature; pixels of neutral color temperature and pixels of warm color temperature

Properties of colors such as intensity and saturation influence the perceptual appearance of color temperature. To take this phenomenon into account, we introduce a temperature strength parameter for fundamental hues that varies from 0 for black to 1 for white.

Figure 6.2 demonstrates the schematic view of the annotation process. It includes several stages. First, segmentation of the image into image blocks/regions. Second, projecting block colors the artistic sphere to extract domain-specific features. Third, annotate image blocks using statistical inference.

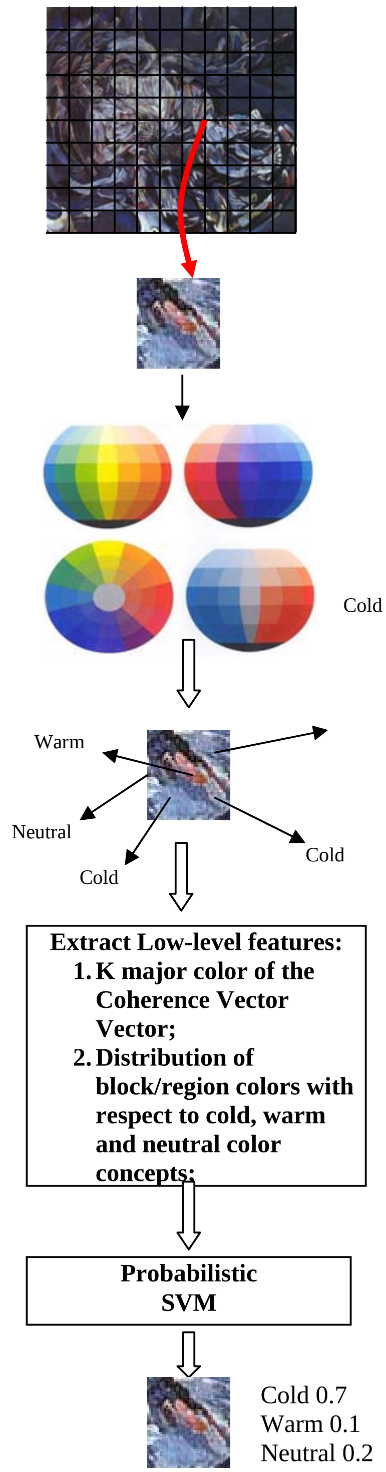


Figure 6. 2 Annotation of color temperature concepts

Next we extract domain-specific features. To minimize the computational complexity, we average the pixel values in a 4x4 neighborhood within each region, calculate the average color in the CIE L*u*v space, and find their corresponding reference color in the artistic color sphere using Equation 6.1. Next, we calculate the spatial coherence of each color temperature within color regions using a modification of the color coherence vector [Pass et al., 1996]. Overall, the feature vector includes the size and average temperature strength of coherent and non-coherent bins, color values of k dominant colors in HSI and CIE L*u*v color spaces and their color temperature extracted from the artistic color sphere. We employ these two color spheres as the color representations in these two color spheres compliment each other and provide more complete information [Herik et al., 2000].

Next, we utilize the calculated feature vector to annotate *warm, cold* and *neutral* color concepts. For this, we employ a supervised machine learning method, the Support Vector Machines (SVM). We use the multi-class probabilistic variant of it developed by Chakrabartty et al. [2002] to generate the posterior probabilities and using the majority vote strategy to assign the color temperature concept for each region. The generated posterior probabilities weighted by the normalized region area, solidity and eccentricity serve as a basis for image ranking during the retrieval stage. The computational time for this method is presented in Table 8.15.

Similarly to color temperature concepts, spatial distribution of colors within a region influences the overall perception of color palette. To analyze the *primary, complimentary* and *tertiary* color palette, we employ a procedure similar to the annotation of color temperature concepts. The only difference is that we now account for the distribution of *primary, complimentary* and *tertiary* concepts within a region. These concepts are discussed in Section 4.3.1.

6. 3. 4 Color Contrast

Based on Itten's theory, we employ analysis of *complimentary, temperature, light-dark* and *value* color contrasts. We analyze color contrast with respect to each pair of neighboring regions. As discussed in Section 6.3.2, our system represents each region as a set of k dominant colors. To effectively represent *complimentary, temperature, light-dark* and *value* color contrasts between the two sets of dominant colors, we adopt the color-pair technique proposed by Chua et al. [1994]. This technique models two neighboring regions as a set of distinct color pairs based on the dominant colors from each region. Figure 6.3 demonstrates the annotation method for color contrast concepts.

We perform the color contrast analysis between two regions in two steps. First, we measure the strength of contrast between two regions and next we account for geometrical properties of these regions to arrive at the final representation. Since we represent two regions as a set of color pairs, we measure the color contrast strength for each distinct color pair and then average the calculated strength across all pairs. In our task, we treat a color pair as distinctive if it exhibits the strength of respective color contrast higher than a predefined threshold. To measure the strength of color contrast between two colors, we find their corresponding reference colors and study their relative location on the artistic color sphere using four measures of color contrast strength. In accordance to Itten's color theory, we operationalize these measures as follows.

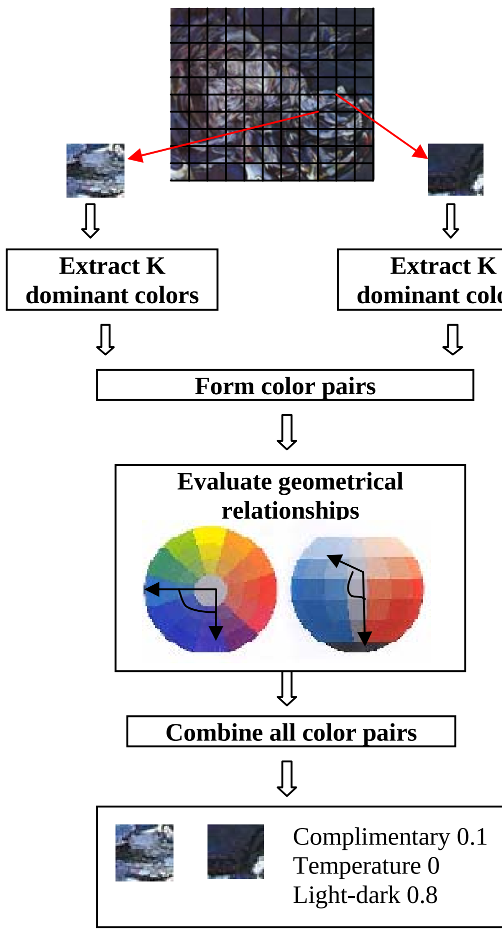


Figure 6. 3 Annotation of color contrast concepts

Value color contrast is defined as the normalized Euclidean distance between the absolute values of the reference color coordinates on the *X* plane of the chromatic color sphere. *Light-dark* and *complimentary* contrasts are defined similarly on the *Z* and *Y* planes of the sphere respectively.

The definition of *temperature* contrast relies on the color temperature concept and the average temperature strength of two neighboring regions. Itten describes *warm-cold* pair as the stronger *temperature* contrast as compared to *neutral-cold* pair. We introduce this heuristics using:

$$w(i,j)=\begin{Bmatrix} 1, if\ t_i='warm' \ \&\ t_j='cold' \\ 0.5, if\ (t_i='warm' \| t_i='cold')\ \&\ t_j='neutral') \\ 0, otherwise \end{Bmatrix} \tag{5. 2}$$

The system calculates color temperature contrast as follows:

$$Ctemp\ (i,j)=w(i,j)*(ts(i)+ts(j))/2 \tag{5. 3}$$

where *ts* denotes the temperature strength of colors *i* and *j* respectively. The temperature strength of each individual color is predefined by Itten's color sphere. To calculate the temperature strength of the overall region/block, we average the temperature strength of block/region colors projected to Itten's sphere. This temperature strength is calculated during previous step discussed in Section 6.3.3.

Next, we annotate each region with color contrast concepts. To facilitate adequate retrieval, we take into account several geometrical region properties, since such parameters as area of the neighboring regions and the length of their border influence human perception of color contrast. For each color contrast concept, we combine the area of two neighboring regions and the length of border between them, weighted by the respective color contrast strength into a normalized sum. This value serves to rank the dataset by image contrast concepts. For example, if the calculated value of a contrast is 0.5 for image regions in the images A and B, but the normalized areas of the regions is 0.5 and 0.1 in the image A and B respectively, then the region in image A will have higher rank when compared to the region in image B.

6. 3. 5 Annotation of Abstract Concepts

In this section, we perform the annotation of abstract concepts *warm, cold, expressive* and *chiaroscuro* as discussed in Section 4.4. The rest of the concepts are not discussed in this thesis since we did not have sufficient datasets for their evaluation. We plan to focus on these concepts in our future work. They are inferred using meta-level artistic concepts of color temperature, color palette and color contrasts. We perform the annotation of abstract concepts using rule-based heuristics described in Table 4.4. To annotate abstract concepts, we perform a three-step procedure. First, we propagate the concept relationships to calculate what visual-level concepts are associated with each abstract-level concept. Next, we extract the values of the respective visual-level color concepts annotated to image regions. Lastly, we average the numerical values associated with each concept to calculate the overall image score with respect to the abstract-level concepts. The annotation of the other abstract-level concepts from Table 4.4 follows the same scheme.

6. 4 Experiment Results

In this set of experiments we evaluate the proposed method for paintings annotation and retrieval based on the artistic color concepts. As we did not require large painting size for color analysis, we rescale the collection to the fixed size of 384x256 or 256x384 similarly to the works of Barnard et al. [2003], Feng et al. [2004] and others. To evaluate the proposed system, we employ the full dataset discussed in Section 5. 3. We employ 315 images to evaluate the color temperature annotation, and utilize the remaining 735 images for testing within the image retrieval framework. As a baseline, we employ simple segmentation method that represents image as nine equal blocks. We refer to the variations of the proposed method based on the Baseline and Blobworld segmentation techniques as Test 1 and Test 2 respectively. We implement Blobworld technique to perform segmentation.

First, we extract 5000 regions from 315 training images. We train the SVM classifier using 25% of the extracted regions and employ 75% to test it. We evaluate the performance of the color temperature annotation using the expert-provided ground truth. The proposed method generates the accuracy of 90% and 85% for Test 1 and Test 2 respectively.

Next, we evaluate the proposed methods within the image retrieval scenario. Using the proposed annotation methods, we label the 735 images with the concepts of color temperature, color palette and color contrast. The retrieval system combines artistic color concepts and geometrical features of the regions to index images. Similar to the first experiment, we utilize the expert-provided ground truth to evaluate the retrieval results. The experts pre-compile ground truth for a variety of queries in four query groups. Each group contains 3 to 5 queries. Table 6.1 demonstrates examples of such queries.

No	Query	Group
1	*Painting in [temp] colors.*	1
2	*Painting in [temp] colors of [palette] palette*	1
3	*Painting with [contrast]*	2
5	*Painting with [temp] region at the [location]*	3
6	*Painting with[contrast] at the [location]*	4
7	*Color of [palette] at the[location]*	3

Table 6. 1. Examples of queries

In this Table, [temp] = {warm, cold, neutral}, [palette] = {primary, complimentary, tertiary}, [contrast] = {complimentary, light-dark, temperature} and [location] = {top, bottom, left, right, centre}. Group 1 covers queries with abstract concepts of color temperature and color palette. Group 2 represents queries with abstract color contrast concepts, while Group 3 incorporates spatial queries of the color temperature and color palette concepts. Lastly, Group 4 represents spatial queries of the color contrast concepts.

Further, we compared the two segmentation methods to highlight the importance of the geometrical information for the retrieval task based on the artistic concepts. For color temperature and color palette queries, the system takes into account the region area, location, solidity and eccentricity. For queries with artistic color contrast concepts, the system considers the area and solidity of regions, location and length of the contacting border.
Table 6. 2 shows the system performance based on Mean Average Precision (MAP) metrics that facilitates the comparison of queries with variable ground truth size.

	Group1	Group2	Group3	Group4
Test 1	0.752	0.402	0.605	0.317
Test 2	0.764	0.720	0.680	0.674

Table 6. 2. Evaluation of the system performance

Overall, the system achieves satisfactory performance for all query groups based on Test 2 segmentation. Test 1 and Test 2 do not differ significantly for Group1 and Group 3 queries, since they do not require elaborate spatial information about the image regions. The difference in the relative performance of Test 1 and Test 2 is most apparent in Group2 and Group4, since these query groups require appropriate information about spatial color distribution. The MAP of Test 2 across all queries is 0.73. Figure 6.4 shows examples of the top images retrieved by the developed system.

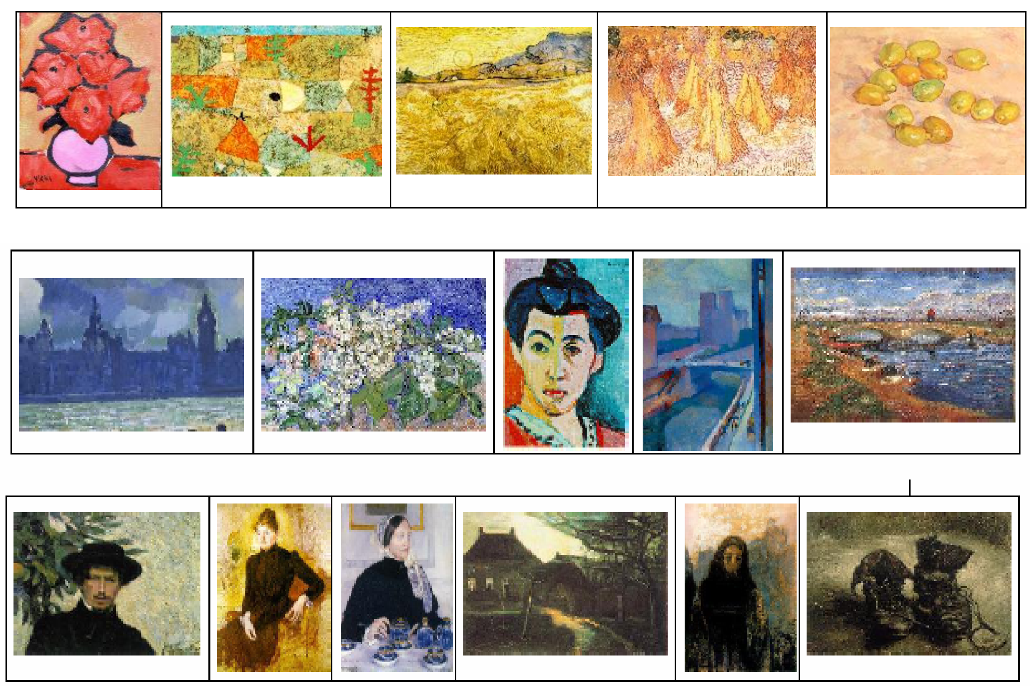


Figure 6. 4. Examples of retrieved images

It demonstrates (in rows from top to bottom): images retrieved by “Paintings in warm colors”, “Paintings in cold colors” and “Paintings with chiaroscuro contrast”.

The existing works [Corridoni et al., 1999; Lay et al., 2004] similarly utilized their proposed indexing methods based on the image retrieval setting. However, they do not evaluate the retrieval performance based on the expert-provided ground truth. Both works reported 100% of syntactic accuracy.

6. 5 Summary

In this chapter, we proposed an automated approach that utilizes domain knowledge of arts domain to analyze and retrieve paintings with color concepts. We performed annotation of major artistic color concepts such as color temperature, color contrast and color palette. These concepts serve as semantic vocabulary for paintings retrieval and provide important cues for auto-annotation of paintings with high-level concepts of artist name, painting style, period of art and culture. The proposed methods utilize spatial information of region colors, which facilitates accounting for a variety of painting styles and extends existing works to handle annotation of paintings in Modern and Contemporary art periods.

Further, we demonstrated the annotation of abstract-level concepts that are widely used among art experts. To index images with the abstract-level concepts, we employed propagation of the concept relationships in the three-level concept ontology. Using this fairly simple annotation method, we demonstrated that accounting of domain-specific knowledge facilitates satisfactory annotation accuracy of abstract-level concepts. However, there are several challenges in the annotation of abstract-level concepts. First, there is a need to experiment with more sophisticated methods for annotation of abstract concepts. Second, as we demonstrated in Section 4.4, abstract-concepts represent a large vocabulary of annotation concepts. To our knowledge, the annotation of these concepts has not been studied yet. Annotation of paintings with these concepts has two benefits. First, it extends the concept vocabulary to handle the expert user needs. Second, it uncovers the semiotic content in paintings due to the fact that artistic theories associate meta-level visual concepts with a variety of symbolic information. For example, color temperatures and contrasts are related to mood in Itten’s theory [1961]. Using this information, it is possible to access additional layers of information available in paintings. Such analysis, however, is beyond the scope of this thesis.

Our experiments in painting retrieval demonstrated that the methods for annotation of meta-level color concepts are effective. In the next chapter, we will discuss methods for annotation of meta-level brushwork concepts.

Chapter 7

Transductive Inference of Serial Multiple Experts for Brushwork Annotation

7. 1 Introduction and Motivation

In the previous chapter, we discussed annotation of visual-level color concepts. To perform annotation of these concepts, we used inductive inference paradigm based on the probabilistic multi-category SVM method to model concepts. This approach assumes that labeled data within each category is consistent as well as the number of labeled samples is sufficient. These assumptions do not always hold for other artistic concepts. In the case of brushwork concepts, each class exhibits a variety of patterns and gathering sufficient labeled data is difficult. Several methods attempted to model the brushwork indirectly to achieve annotation of artist name concepts [Herik et al., 2000; Li et al., 2004]. In our work we implement similar approach and utilize it as a baseline. These methods directly model the artist profile based on low-level texture features. Such an approach has several drawbacks. First, it does not incorporate domain-specific knowledge for the disambiguation of results. Second, since brushwork is not represented explicitly in such a framework, the introduction of other high-level concepts in arts domain will require additional training. In Chapter 5 we proposed the framework for ontology-based annotation, which utilizes the meta-level artistic brushwork concepts within the annotation process. This framework alleviates the problems of traditional statistical learning by the use of domain-specific ontology. In this chapter, we focus on the annotation of brushwork patches with artistic brushwork concepts. To our knowledge, this is the first attempt to explicitly model artistic brushwork concepts for the purpose of advancing the ontology-based annotation in the paintings domain. To address the problem of effective annotation with brushwork concepts, we need to tackle three challenges.

First, we utilize a number of statistical and signal processing features for the representation of brushwork contents for adequate representation of a large variety of brushwork patterns. This yields high-dimensionality of the feature space, leading to the 'curse of dimensionality'. It essentially means that the sparseness of data increases exponentially with the dimensionality of the input space given a constant amount of data, with points tending to become equidistant from one another at a certain high dimension [Friedman, 1994]. This phenomenon largely degrades the quality of the traditional inference methods and poses the need for feature selection methods.

Second, we need to explore techniques for image annotation based on a small set of labeled patterns. Since manual annotation of art images is very tedious and costly, usually only limited datasets are available to perform the classifier training. Similar to the existing studies in the paintings domain [Herik, et al., 2000; Breen et al., 2002; Li et al., 2004], we perform block-level analysis for the brushwork annotation that results in a large amount of unlabeled data. We aim to construct more accurate classifiers based on the combination of labeled and unlabelled data. We reviewed these methods in Section 3.6.3.

Third, a vast variety of brushwork patterns poses the need for robust classifiers. The data mining community and related communities have devoted much effort to develop techniques for creating better classifiers [Barnard et al., 2003; Murphy et al., 2003; Skounakis et al., 2003; Gyftodimos et al., 2004] and, more recently, combining individual classifiers to produce a more accurate combined classifier [Kuncheva, 2004].

7. 2 Related Work

Early work on expert combination mostly focused around 'multiple experts vs multiple levels' comparisons, where the authors were concerned with the structure of decision hierarchies [Gluskman, 1971; Schueermann, 1983]. Recent studies have shown that the use of multiple expert approaches could lead to higher accuracy when compared to the single classifier approach [Kittler et al, 1996; Pudil et al., 1992].

There are several benefits of the multiple classifiers (or experts) approach. First, it partitions the problems and decreases the complexity of probability estimation. Second, since several independent classifiers contribute to the overall decision, this approach requires smaller training sets as compared to hierarchical learning approaches [Barnard et al., 2003; Gyftodimos et al., 2004; Murphy et al., 2003]. Third, multiple expert frameworks facilitate dimensionality reduction of the feature sets, since the overall classification task is composed of several focused sub-tasks. Lastly, the modular organization of sub-tasks facilitates the incorporation of domain knowledge especially into their inter-dependence and interaction with the target function.

Rahman et al. [1999; 2000] discussed a generic approach to combinations of multiple experts. Configurations that combine experts in several sequential levels are called *serial* combinations [Pudil et al., 1992]. The main attraction of the serial approaches is that these configurations: 1) implement a step-wise disambiguation of patterns and 2) facilitate reduction of rejection rate. In statistical pattern recognition, the reject option has been introduced to guide the classifier against excessive errors [Devijner et al, 1982]. If the rejection option is allowed, the quality of recognition increases, but on the whole fewer patterns are recognized. In the serial combinations framework, the number of rejected patterns is minimized due to the re-evaluation of ambiguous patterns in subsequent levels of individual experts.

Individual experts facilitate the use of both inductive and transductive inference to generate their decisions. We discussed the relationship between inductive and transductive inference in Section 2.4. Recently, many studies focused on transductive inference for annotation of large data collections due to its applicability to many real-world situations. A non-exhaustive list of recent contributions includes [Vapnik, 1982, 1998; Joachims, 1999; Demiriz et al., 2000; Wu et al. 1999; Blum et al., 2003; Debreko et al, 2004, El-Yaniv et al., 2004]. The works of Joachims [1999], Demiriz et al.[2000], Wu et al. [1999] and El-Yaniv et al.[2004] dealt with algorithmic issues, while Vapnik [1982, 1998], Blum et al. [2003] and Debreko et al. [2004] focused on the theoretical discussion and performance bounds. Vapnik's [1982, 1998] and Blum et al. [2003] offered the formulation for implicit bounds. Explicit PAC-Bayesian bound was presented in work of Debreko et al. [2004]. El-Yaniv et al.[2004] proposed a transductive learning scheme based on this bound. This method yields comparable results with TSVM proposed by Joachim [1999] for image classification task.

In accordance to [Rahman et al., 2000], the use of relevant features minimizes the error propagation through the framework. A large body of studies [Blum et al, 1997; Kohavi et al., 1997] has proposed techniques for dimensionality reduction. The well-known approaches to dimensionality reduction are *feature selection* and *feature transformation* techniques [Parsons et al., 1994]. Feature selection attempts to discover the most relevant attributes. It includes wrapper approaches and filter approaches. In wrapper approaches [Kohavi et al., 1997], the relevant feature subset is induced from error rates of the classifier. In filter approaches, the measure of feature subset quality is independent of classifiers; it is based on its correlation with the target function.

Feature transformation techniques such as Principle Component Analysis (PCA) transform the original space into a lower dimensional space. PCA is not always applicable since the variance is not necessarily correlated to the discriminative power. Another drawback of methods like PCA is the lack of interpretability of the newly formed feature set.

In the proposed framework, we employ a hybrid of the feature filtering and the feature wrapper approaches, since it first estimates the feature relevance based on the feature value distributions and then iteratively select the most discriminative features based on the classification accuracy of the model. The features are scored based on the Chi-square statistics. This so-called *symmetric* method measures the association between the two distributions [Lehmann, 1999]. Chi-square statistics is used to select features for various tasks such as the rule induction task [Imam et al., 1999] and text categorization [Yang et al., 1997].

7. 3 Brushwork Representation

In Section 4.3.2, we introduced domain-specific knowledge about brushwork classes. We examined various properties of brushwork and justified the use of the texture-based approach for brushwork analysis. In this section, we focus on the low-level features used to represent brushwork classes. Table 7.1 provides the summary of brushwork classes. It contains examples of brushwork patterns in each of the analyzed classes and the relevant features for each class.

Various comparative studies showed that no single texture features representation approach performs best for all kinds of textures. Hence, to capture the variety of patterns in our dataset, we utilize various signal-based and statistical texture feature representations. As Table 7.1 demonstrates, our collection includes a vast number of patterns, which are mostly stochastic.

They exhibit a variety of properties such as directional (for example, *impasto*), non-directional (*pointillism*), contrasting (*divisionism*) and smooth (*mezzapasta*). In terms of the spatial homogeneity we can roughly group brushwork patterns as homogeneous (*mezzapasta* and *pointillism*), weakly homogeneous (*divisionism*) and non-homogeneous (*scumbling*, *shading* and *glazing*). We utilize color and texture features for pattern representation. To calculate color features, we utilize the CIE L*u*v color space. From color histograms, we extract major colors with account for their perceptual similarity [Chua et al., 1994]. We calculate complimentary and chiaroscuro color contrasts based on our previously developed method [Marchenko et al., 2005].

In order to model the variety of brushwork patterns, we use several texture features. First, we make use of the edge-based features to capture linear components of a pattern. We apply Canny edge detector [Canny, 1986] with a fixed threshold to the whole collection and calculate the directional histogram:

$$EdgeHist = \frac{P_i}{\sum_i P_i} \tag{7. 1}$$

where P_i denotes the number of edge pixels in the *i-th* direction. Next we extract gradient-based features. These are statistics of image gradients (mean and deviation) and their directional histogram. We calculate the directional gradient histogram using the formula above. For both histograms, we employ eight directions.

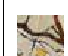
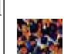
Class	Properties	Low level features	
Shading	Edges and gradients, often directional, intensity contrast, weakly or non-homogeneous	Multiscale Gabor texture features, Zernike moments, Chiaroscuro (intensity) color contrast, Multiscale Fractal Dimension, Lacunarity	
Glazing	Subset of hues (yellow, red, orange), intensity contrast, gradients, non-homogeneous	Top major colors with account for the perceptual similarity, Chiaroscuro (intensity) color contrast, Daubichies Wavelet Transform, Zernike moments, Multiscale Fractal Dimension, Lacunarity	
Mezzapasta	Homogeneous, low intensity contrast and small gradients	Mean and Deviation of image magnitudes, Directional Histogram of Gradient Magnitudes, Major colors with account for perceptual similarity	
Grattage	Edges, high gradients, intensity contrast, inhomogeneous	Number of Edge Pixels, Mean and Deviation of Directional Edge Histogram, Chiaroscuro (intensity) color contrast, Daubichies Wavelet Transform, Multiscale Fractal Dimension, Lacunarity	
Scumbling	Soft gradients, low intensity and hue contrast, low directionality, weakly homogeneous	Daubichies Wavelet Transform, Zernike moments, Chiaroscuro (intensity) and Complimentary (hue) color contrast, Multiscale Fractal Dimension, Lacunarity	
Impasto	Edges, high gradients, often directional, low hue contrast, high intensity contrast	Number of Edge Pixels, Directional Histogram of Gradient Magnitudes, Chiaroscuro (intensity), Complimentary (hue) color contrast, Daubichies Wavelet Transform, Multiscale Gabor texture features	
Pointillism	Medium intensity contrast, medium roughness, no directionality, homogeneous	Mean and Deviation of Magnitude, Chiaroscuro (intensity) color contrast, Daubichies Wavelet Transform, Zernike Moments	
Divisionism	High gradients, high roughness, high intensity and hue contrast, no directionality, weakly homogeneous	Mean and Deviation of Magnitude, Daubechies Wavelet Transform , Chiaroscuro (intensity) and Complimentary (hue) color contrast, Multiscale Fractal Dimension, Lacunarity	

Table 7. 1. Low-level features for the representation of brushwork classes

For representing the directional characteristics, we utilize multi-scale Gabor Transform proposed for image retrieval by Majunath et al. [1996]. A Gabor filter bank is a pseudo-wavelet filter bank where each filter generates a near-independent estimate of the local frequency content. Gabor filter acts as a local band-pass filter with certain optimal joint localization properties in the spatial domain and spatial frequency domain. To extract Gabor features, the input image *I(x, y)* is convolved with a set of Gabor filters of different orientations and spatial frequencies that cover appropriately the spatial frequency domain. In our experiments, we utilize 8 orientations and 4 scales. The general functional *g(x,y)* of the two-dimensional Gabor filter family can be represented as a Gaussian function modulated by an oriented complex sinusoidal signal:

$$
\begin{aligned}
g(x,y) &= \frac{1}{2\pi\sigma_x\sigma_y}\exp[-\frac{1}{2}(\frac{\tilde{x}^2}{\sigma_x{}^2}+\frac{\tilde{y}^2}{\sigma_y{}^2})+2\pi jW_{\tilde{x}}] \\
\tilde{x} &= a^{-m}(x\cos\theta+y\sin\theta),\ \tilde{y} = a^{-m}(-x\sin\theta+y\cos\theta)
\end{aligned}
\qquad (7.\ 2)
$$

where σ denotes the parameters of the filter with respect to x and y, W is the center frequency, and θ determines the orientation of the filter, a^{-m} is the scale factor to ensure that the energy is independent of scale m.

Another important texture feature is the Dyadic Wavelet Transform (DWT). DWT is most useful for multi-resolution image analysis and captures a variety of texture properties [Mallat, 1989]. Dyadic wavelet decomposition is carried out using 2 channel filter banks composed of a low-pass and a high-pass filter and each filter bank is sampled at a half rate (1/2 down sampling) of the previous frequency. We employ Daubechies filter banks for our study. This filter bank has the important qualities of orthogonality and compact support.

To extract texture features from Gabor and Daubechies filter response, we calculate the mean and deviation of energy distribution of the transform coefficients for each sub-band at each decomposition level. Let the image sub-band of size NxN be $I_i(x, y)$ with i denoting the specific sub-band, then the resulting feature vector obtained from the filter response is $\{\mu_i,\sigma_i\}$ with,

$$
\begin{aligned}
\mu_i &= \frac{1}{N^2}\sum_{x=1}^{N}\sum_{y=1}^{N}\left|I_i(x,y)\right| \\
\sigma^2{}_i &= \frac{1}{N^2}\sum_{x=1}^{N}\sum_{y=1}^{N}\left|I_i(x,y)-\mu_i\right|^2
\end{aligned}
\qquad (7.\ 3)
$$

The major drawback of energy-based features above is the implicit assumption of texture homogeneity. Such assumption does not hold for several classes of brushwork in our dataset that exhibit non-regular textures (for example, *scumbling* and *shading*).

To represent non-regular textures, Mandelbrot [1982] popularized the self-similar fractional Brownian motion (fBm) model, which is characterized by a single parameter known as the *Hurst parameter*. The Hurst parameter controls the visual roughness of the process at all scales. In our study, we utilize the extended self-similar (ESS) model [Kaplan et al., 1995] that measures the Hurst Parameter at various scales and, thus, encodes more detailed textural information. First, the ESS model calculates the directed increments (in x and y orientation) of dyadic scales for an image $I(x,y)$:

$$\begin{aligned} \Delta_s^{Xaxis}(x,y) &= I(x+2^s, y) - I(x,y) \\ \Delta_s^{Yaxis}(x,y) &= I(x, y+2^s) - I(x,y) \end{aligned} \tag{7. 4}$$

The structure function is defined as the average of the incremental power over all available pixels:

$$f_s^{\theta} = \frac{1}{N(N-2^s)} \sum_x \sum_y |\Delta_s^{\theta}(x,y)|^2 \tag{7. 5}$$

for $\theta = \{X_{axis}, Y_{axis}\}$. The multi-scale Hurst parameters are computed for scale s to obtain the isotropic and directed features as follows:

$$\begin{aligned} H_s(x,y) &= \frac{1}{2}\log_2\left(\frac{f_{s+1}^{Xaxis} + f_{s+1}^{Yaxis}}{f_s^{Xaxis} + f_s^{Yaxis}}\right) \\ H_s^{\theta}(x,y) &= \frac{1}{2}\log_2\left(\frac{f_{s+1}^{\theta}}{f_s^{\theta}}\right) \end{aligned} \tag{7. 6}$$

Finally, we utilize statistical moment descriptors to extract the surface information from the brushwork patches. We employ these features to represent *glazing, shading* and *scumbling* classes. Teague [1979] first introduced the use of Zernike moments to overcome the shortcomings of information redundancy present in the popular geometric moments. Zernike moments have the property of orthogonality and have been shown to be effective in terms of the image representation. Zhang et al. [2001] demonstrated that Zernike moments out-perform geometrical moments in shape retrieval task. Another important property of Zernike moments is that they are rotation invariant and can be easily constructed to an arbitrary order. The Zernike polynomials are a set of complex, orthogonal polynomials defined over the interior of a unit circle $x^2+y^2=1$ as:

$$\begin{aligned} V_{nm}(x,y) &= V_{nm}(r,\theta) = R_{nm}(r)e^{jn\theta} \\ R_{nm}(r) &= \sum_{s=0}^{\frac{n-|m|}{2}} (-1)^s \frac{(n-s)!}{s!\left(\frac{n+|m|}{2}-s\right)!\left(\frac{n-|m|}{2}-s\right)!} r^{n-2s} \end{aligned} \tag{7. 7}$$

where n is non-negative integer, m is the number such that n-$|m|$ is even and $m \leq n$, $r=sqrt(x^2+y^2)$ and $\theta=tan^{-1}(x/y)$. The magnitude of Zernike moments has the property of rotational invariance and is defined as:

$$A_{nm} = \left| \frac{n+1}{\pi} \sum_x \sum_y I(x,y) * V_{nm}(x,y) \right| \tag{7. 8}$$

where $x^2+y^2 \leq 1$ and * denotes the complex conjugate. For our task, we calculate 32 Zernike moments.

We employ all of the above features for adequate representation of brushwork patterns. This yields high-dimensionality of the feature space. However, only a subset of features is relevant to individual brushwork classes. To tackle this problem, we adopt an approach that combines several experts, each of which assigns candidate classes to the unlabelled patterns based on a subset of features. In the next session, we briefly discuss a generic framework for serial combination of multiple experts.

7. 4 Generic Multiple Serial Expert Framework for Annotation

The decision process of the multiple serial expert framework is pre-defined by the decision hierarchy, which encodes the sub-tasks and relationships among them. Each level of the decision hierarchy includes several individual experts that operate simultaneously and independently of each other. We represent the decision hierarchy as the decision tree that consists of a root-node, a number of non-terminal nodes and a number of terminal nodes. Associated with the root node is the entire set of classes into which a pattern may be classified. A non-terminal node represents an intermediate decision and its immediate descendant nodes represent the decisions originating from that particular node. After the first intermediate decisions are taken at the preliminary level in the decision hierarchy, the final decision is reached through a step-wise refinement procedure. As the decision hierarchy is traversed in the forward direction, the decisions of individual experts become more and more refined, and the confidence associated with the decision increases. The decision making process terminates at a terminal node, where the unlabelled patterns receive their respective labels. Figure 7.1 demonstrates the decision hierarchy that incorporates these ideas.

The aim for the decision hierarchy is to reduce the target set or the subset of classes to which a pattern might belong. Individual experts, which are associated with the decision tree nodes, perform such reduction. We formalize the reduction of target size as follows. The expert at the *i-th* level receives the input vector (X,S_{i-1}), where X represents a pattern and S_{i-1} denotes the decision of the ancestor node. This expert generates its own decision S_i, which essentially represents a set of classes to which the pattern X belongs with the maximum confidence. The set S_i is a subset of its respective set S_{i-1} $(S_n \subset S_{n-1} \subset S_i \ldots \subset S_0)$. When pattern X reaches the terminal node, it is labeled with a single element of S_i.

There are several important issues regarding the multi-expert frameworks. First, since the serial expert approach sequentially refines its decisions, then the multi-expert configuration cannot exceed the performance of its terminal nodes, provided that all experts operate on the same feature space and dataset. Therefore, the final performance can be either lower or identical to the performance of terminal nodes if all the experts utilize the same feature. However, if the experts operate on their respective relevant feature subsets, then the sparseness and noise of the feature space from the point of view of the expert are reduced and the overall accuracy of the combined system is expected to be satisfactory [Rahman et al., 1999].

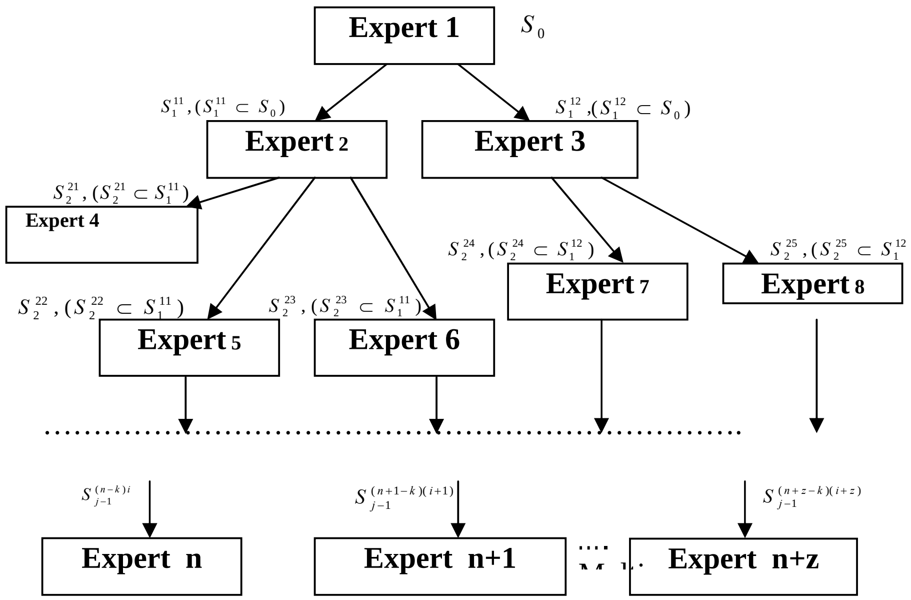


Figure 7. 1. Serial Combination of Multiple Experts

Second important issue in the design of the decision tree. The order of the sub-tasks influences the overall performance, since the performance of the subsequent levels of experts depends on the performance of the pervious levels. The number of levels should be optimal such that the increase of performance achieved by incremental enhancement does not diminish as more experts are combined. As argued by Rahman et al. [1999], the design of decision tree relies on the knowledge about the classification task. In Section 7.5.1, we define the decision hierarchy used in annotation of brushwork patterns.

Third, Rahman et al [1999] outlined two major strategies for annotation using the serial multi-expert approaches: Class Set Reduction and Class Reevaluation. We will discuss these strategies in the next two sections and evaluate the proposed multi-expert framework with respect to both strategies in Section 7.8.

7. 4. 1 Class Set Reduction strategy

The Class Set Reduction strategy requires that the experts evaluate all samples from the ancestor node and pass them to the subsequent experts. There are three sources of information for any expert. First, the current unlabelled patterns. Second, the list of candidate classes passed on by the ancestor expert. Third, the desirable subsets of candidate labels to be generated from this list. The candidate class labels reflect the choice of the previous expert in identifying the current set of unlabelled patterns. Thus, the expert at i-th level of the decision hierarchy needs to produce a candidate class subset S_i of its own preferences as a function of each unlabelled pattern X. The subset S_i should have a high probability of containing a true label among the candidate class labels corresponding to the pattern. Here we present the formulation of this annotation strategy proposed by Rahman et al. [1999]. Assuming that:

- $w(X)$ is the true class of pattern X,
- $d(X,S_i)$ is the candidate class generated by the current expert,
- P_{ei} is the probability that S_i does not contain true class, $P_{ei}=P[w(X) \notin S_i]$,
- P_{ci} is the probability that S_i contains true class, $P_{ci}=P[w(X) \in S_i]$,
- $P_{e(i+1)}$ is the probability that the expert at $(i+1)$ level assigns X to the wrong class, although S_i contains the true class label $P_{e(i+1)}= P[d(X,S_i) \neq w(X) \mid w(X) \in S_i)]$,
- $P_{c(i+1)}$the probability that the expert at $(i + 1)$ level assigns X to the correct class, given that S_i contains true class index, $P_{c(i+1)}= P[d(X,S_i) = w(X) \mid w(X) \in S_i)]$.

Then the overall correct classification of n-level serial network is

$$P_{cT} = P_{c1} \times P_{c2} \times \ldots \times P_{cn} \tag{7. 9}$$

and the overall error of n-level serial network is

$$\begin{aligned} P_{eT} &= (P_{e2} + P_{e1} \times P_{c2}) + (P_{e3} + P_{c1} \times P_{e2} \times P_{c3}) + \\ &+ \ldots + (P_{en} + P_{c1} \times P_{c2} \times P_{c3} \times P_{e(n-1)} \times P_{cn}) \end{aligned} \tag{7. 10}$$

Here, since each unlabelled pattern is evaluated until it reaches the leaf nodes, the probability of the correct and erroneous labeling depends on the outcome of the preceding levels. In the Class Set Reduction strategy, the ability to pass samples to the next level is important, since it increases the chance of an unlabelled pattern being assigned the true label. Thus, it assumes zero rejection rate at the intermediate nodes.

7. 4. 2 Class Reevaluation strategy

In contrast to the Class Set Reduction strategy, the Class Reevaluation does not require the experts to pass all instances to the subsequent levels. It extends the intermediate nodes to facilitate additional analysis: if the unlabelled patterns are assigned labels with high confidence, then these assignments become final and the decision process does not evaluate these patterns further. In essence, this strategy reevaluates patterns that are assigned with the confidence lower than some predefined threshold (t_{accept}). Such strategy requires the individual experts to perform recognition with respect to individual classes, and pass the patterns with ambiguous assignments to the next level.

We now formalize the decision process for unlabelled pattern X. Assuming:

- $w(X)$ is the original class associated with the current pattern,
- $d(X, t_{accept})$ denotes the candidate class of pattern X generated by the current expert based on the confidence threshold
- α denotes the confidence of expert in assigning a candidate class to pattern X,
- P_{ci} is the probability that the expert generates the true class,
  $P_{ci} = P[d(X,t_{accept}) = w(X)]$,
- P_{ei} is the probability that expert doesn't generate true class.
  We define $P_{ei} = P_{error} + P_{rejection}$, where
- $P_{error} = P[d(X,t_{accept}) \neq w(X) \mid (\alpha > t_{accept})]$ denotes the probability of erroneous class label assigned to the unlabelled pattern X with confidence α higher then threshold t_{accept}, and
- $P_{rejection} = P[d(X,t_{accept}) = w(X) \mid (\alpha < t_{accept})]$ denotes the probability of the correct class label assigned and being rejected due to the confidence lower then the threshold,

Similarly to the Class Reduction strategy, the probability of correct decision is defined as:

$$P_{cT} = P_{c1} \times P_{c2} \times \ldots \times P_{cn} \tag{7. 11}$$

with the errors given by $P_{eT}=1-P_{cT}$.

7. 5 Transductive Inference of Brushwork Concepts Using Multiple Serial Experts Framework

In this section we discuss a proposed multi-expert framework that employs transductive inference of brushwork concept annotation. Figure 7.2 demonstrates the framework for transductive inference of brushwork patterns.

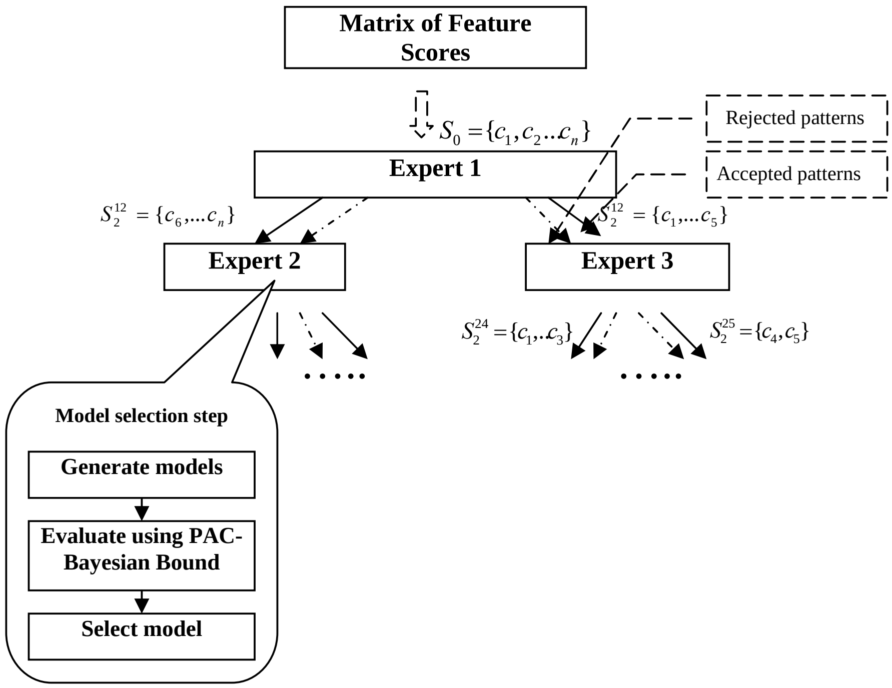


Figure 7. 2. Serial Combination of Multiple Experts

At the pre-processing stage, the system performs scoring of features that measures their discriminative power with respect to the brushwork classes. As the decision process traverses the decision hierarchy, it selects the most discriminative features for individual experts based on their respective sub-tasks. The proposed framework implements concept inference at the local and global levels.

At the local level, individual experts implement transductive inference scheme proposed by El-Yaniv et al. [2004]. We will discuss this inference in Section 7.6. As a result, an individual expert generates a cluster space C with k clusters $\{C_i\}$ for $i = 1 \ldots k$. Clusters include both labeled and unlabelled patterns $C_i = N_l \cup N_u$, where N_l and N_u denote labeled and unlabeled patterns respectively. The expert performs annotation of the unlabelled patterns using the cluster purity measure. We define pure cluster of class X as the cluster in which more than 75% of the labeled patterns belong to the target sets. The decision tree pre-defines the target sets for each individual expert. In essence, the target sets represent pair-wise constraints "can" and "can not", specifying which labels can be grouped together within a cluster. The cluster purity represents the degree to which the calculated cluster c contains labels of the target set X and is defined as:

$$p(c,X) = N_X / N_{all} \tag{7.12}$$

where N_X and N_{all} denote the number of labeled patterns of class X and the overall number of patterns in cluster c respectively. Thus we view the resulting cluster space as follows:

$$C = C^p \cup C^{np} \cup C^{nl} \tag{7.13}$$

where C^{nl} denotes clusters that include unlabelled samples only and, thus, carry no class information (labels), C^{np} and C^{p} denote clusters represent a mixture of labeled and unlabeled samples. The C^{p} clusters are pure; the unlabelled patterns in these clusters receive their candidate labels. The unlabelled samples in clusters C^{nl} and C^{np} are rejected.

At the global level, the framework performs inference that estimates the candidate classes for the rejected samples based on the decision hierarchy. The global inference mechanism passes the unlabelled samples, which are rejected at the current level, to all experts at the next level. These experts re-evaluate rejected samples and, based on their decisions, either accept them or reject them again. This process of re-evaluation continues at the next level of the decision hierarchy and so on, until the samples are either accepted or reach the leaf nodes. If the samples are rejected at the level proceeding the leaf nodes of the decision hierarchy, the global inference mechanism forces their evaluation in every expert of this level and assigns the candidate a label based on the highest confidence value generated by these individual experts. If the patterns are rejected everywhere, we assign them to the most probable label in the subset of candidate labels that preceded its rejection.

The local and global inference mechanisms facilitate both Class Reduction and Class Reevaluation strategies discussed in Sections 7.4.1 and 7.4.2. The implementation of the Class Reduction strategy is straightforward due to the global inference mechanism. The Class Reevaluation strategy relies on the local inference mechanism. In Formula 7.13 we defined the cluster space in terms of pure and impure clusters. The pure clusters C^{p} further include the clusters that contain a majority of samples labeled with a single class X_i. If the purity measure of these clusters exceeds the pre-defined threshold t_{accept}, the decision process assigns the final labels to the unlabelled samples in these clusters in accordance to the Class Reevaluation strategy.

7. 5. 1 Decision hierarchy

In our task we know *apriori* the characteristics of the brushwork classes. We rely on such characteristics to formulate the sub-goals at the intermediate and terminal nodes.

Rahman et al. [1999] demonstrated experimentally that two-level configurations produce very good results. In our study, we employ the three-level decision hierarchy with the single brushwork class corresponding to the terminal node. Figure 7.3 represents the decision hierarchy for the brushwork annotation.

The decision process starts with all classes and the original dataset. At the first level, we arrange the brushwork classes in the subsets based on the degree to which they exhibit similar linear components. We define the three sub-goals as {*impasto, grattage* and *divisionism*}, followed by {*scumbling, glazing* and *shading*} and, finally, {*pointillism* and *mezzapasta*}.

The brushwork classes *impasto, grattage* and *divisionism* exhibit strong linear components, while are non-homogeneous textures with soft gradients and linear components and classes pointillism and mezzapasta are homogeneous patterns without linear components.

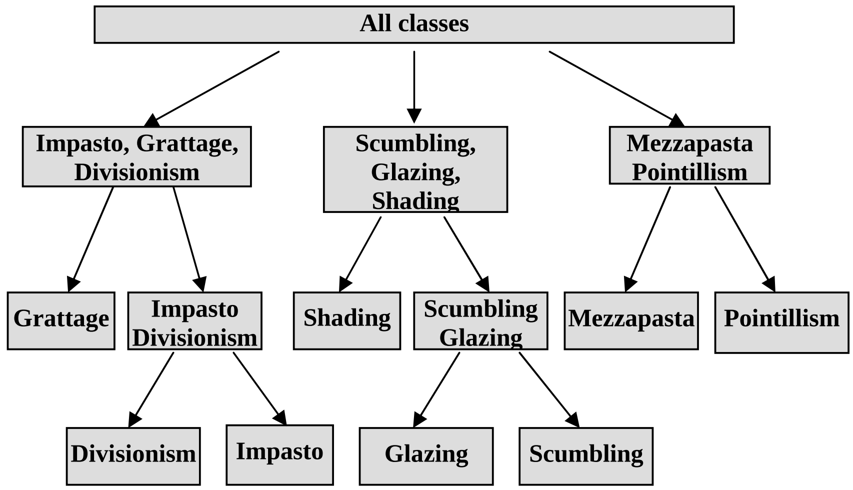


Figure 7. 3. The decision hierarchy for brushwork annotation

At the next level, there are three experts working simultaneously on their respective datasets. The first second-level expert aims to split *grattage* class from *impasto* and *divisionism* classes, since patterns in *grattage* class exhibit long edges and high *chiaroscuro* contrasts compared to the other two classes. The second expert assesses its input patterns by roughness. This leads to the terminal node *shading*, since this class exhibits more roughness as compared to *scumbling* and *glazing*. The third expert analyzes the patterns belonging to only two classes, and hence produces the terminal nodes for *mezzapasta* and *pointillism* since these classes vary with respect to the roughness and the number of colors they exhibit.

7. 5. 2 Feature Selection

The major aim of feature selection task is to provide individual experts with the feature set relevant to their respective sub-task. The multi-expert framework supports both manual and automatic selection of features.

7. 5. 2(a) Manual Feature Selection

Using apriori knowledge about the brushwork classes from Table 7.1, we assign relevant features to individual experts. The details of the image features we use are discussed in Section 7.3. Figure 7.4 demonstrates the decision hierarchy of individual experts with their respective relevant features.

7. 5. 2(b) Automatic Feature Selection

To avoid manual assignment of features, we developed a method for automatic feature selection based on the statistical properties of the feature distribution. This method calculates the feature discriminative scores with respect to the brushwork classes using a three-step procedure.

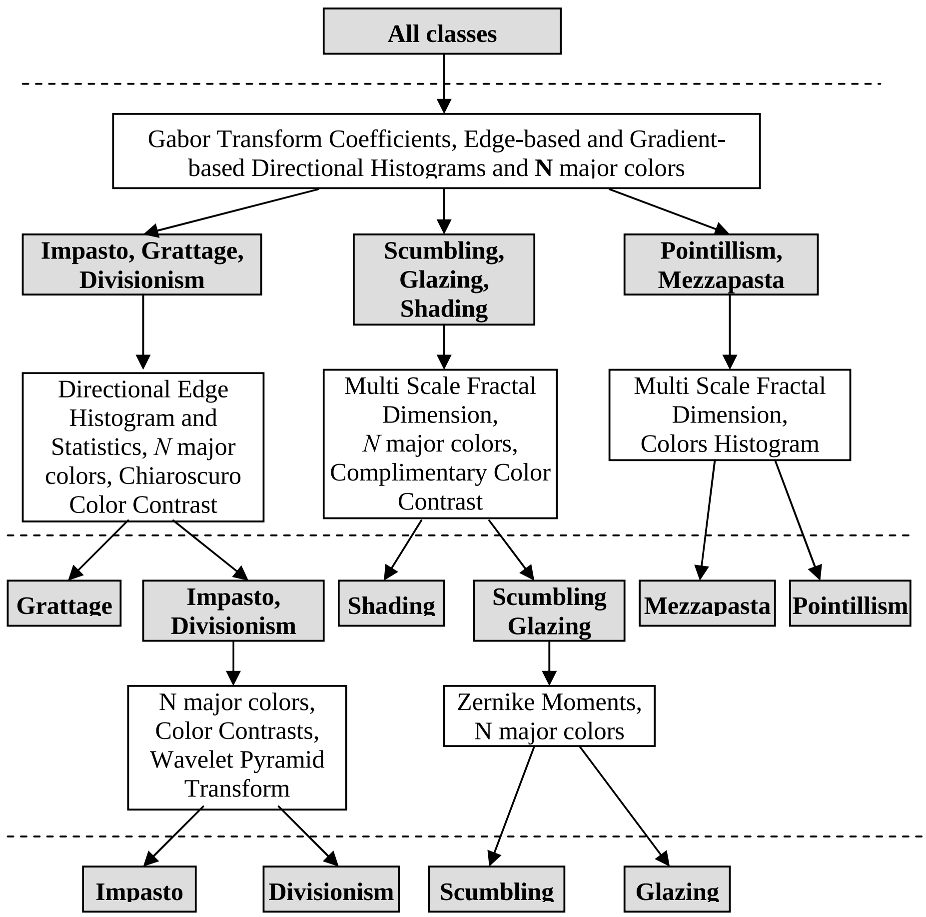


Figure 7. 4. The decision hierarchy for brushwork annotation

First, it calculates tight clusters in the feature space using iterative K-means method. Since the K-means clustering method minimizes the intra-cluster distance, the data points within a cluster are somewhat close to each other in the feature space and exhibit relatively small variances along some of the feature dimensions. Thus, feature dimension is more likely to be relevant to the cluster if the projection of the cluster onto this dimension has a smaller variance.

Second, the proposed method assesses the "importance" of feature for the calculated clusters. For this, it employs Pearson's Chi-square statistics that facilitates measurement of "goodness-of-fit" between observed and expected distributions. Low "goodness-of-fit" value signifies that distributions are different, while high value indicates their similarity. To score the features, we treat distribution within a cluster as observed and distribution of the whole dataset as expected distribution. Intuitively, if the distributions of feature in a cluster and in the whole dataset are similar, then the analyzed feature is not representative of the cluster and its Chi-square statistics is comparatively low. To represent feature distribution, we employ normalized histograms $O=\{O_1, O_2, \ldots, O_{100}\}$ and $E=\{E_1, E_2, \ldots, E_{100}\}$ for the observed and expected distributions respectively. We measure "goodness-of-fit" using the following formulae:

$$\chi_j^2 = \sum_{i=1}^{100} (O_{ij} - E_i)^2 / E_i \tag{7. 14}$$

where the counts O_i and E_i denote *i-th* histogram bin count of the observed and expected feature distributions respectively, χ^2 denotes the discriminative score of *j-th* feature with respect to a currently analyzed cluster. Lastly, the method combines the feature scores of clusters to achieve the score of *j-th* feature of the brushwork classes using the following formula:

$$\mathrm{S}core(X, j) = \sum_{c=1}^{K} p(c, X) \times A(c) \times \chi_j^2(c) \tag{7. 15}$$

where *p(c,X)* denotes cluster purity of cluster *c* with respect to class *X*. Equation 7.12 demonstrates how we calculate the cluster purity. *K* is the total number of clusters, *A(c)* denotes the size of cluster *c* normalized by the total number of labeled patterns in class *X* and χ^2(c) denotes Chi-square statistics of *j-th* feature in cluster *c.*

To select relevant features to their sub-tasks, the expert utilizes information from matrix of feature scores and the decision tree hierarchy, since, the decision hierarchy pre-defines subset of brushwork classes for the decision tree hierarchy. As Figure 7.4 demonstrates, the decision tree hierarchy predefines the subset of analyzed brushwork classes for each sub-task. Individual experts utilize this information and extract the feature discriminative scores for their respective subsets of classes and further utilize these scores during model selection step as discussed in Section 7.6.3.

7. 6 Individual Experts

To implement individual experts, we employ transductive inference method since they account for distribution of unlabelled samples and possibly lead to more accurate results [Vapnik, 1982]. For simplicity, we rely on the transduction formulation for binary classification proposed by Vapnik [1998]. In this formulation the expert is given a *full sample*

X_{l+u} of l labeled and u unlabelled patterns. Based on the labeled and unlabeled patterns, the expert's goal is to predict, as accurately as possible, the labels of the remaining unlabeled points, which constitute the *test set*, $X_u = X_{l+u} - X_l$.

We denote by H a set of binary hypotheses consisting of functions from the input space X to $Y = \{+/-1\}$. The experts's goal is to choose a good hypothesis from H. For each hypothesis $h \in H$ and a set of samples $Z = x_1 \ldots x_N$ we denote the full sample risk of hypothesis h as follows :

$$R_h(Z) \cong \frac{1}{N}\sum_{i=1}^{N} L(h(x_i), \phi(h_i)) \qquad (7.\ 16)$$

where $\phi(x_i) \in Y$ denotes the true label of pattern x_i and $L(\cdot,\cdot)$ denotes loss. $R_h(X_u)$ is referred to as the *transduction risk* or *test error* (of h), and $R_h(X_l)$ is the *training error* (of h). The goal of the expert is to choose $h \in H$ with minimal *transduction risk* $R_h(X_u)$.

Similar to other studies [Miller et al., 2003; El-Yaniv et al., 2004], we employ transductive learning via clustering for brushwork annotation. This approach is appropriate to our task for several reasons. First, the clustering techniques model a class as a set of clusters (distributions) in the feature space. Second, they incorporate unlabelled patterns and facilitate annotation with relatively small labeled dataset (so called *semi-supervised annotation*). Lastly, in many circumstances the data density can provide good clues regarding what data points belong to what classes. In our work, we employ hierarchical [Murtagh, 1983], k-means [Hartigan 1975; Hartigan et al., 1979] and probabilistic clustering using Gaussian Mixture Models [McLachlan et al., 1988].

7. 6. 1 Transductive Risk Estimation

Several bounds were proposed for transductive risk estimation. In this study we employ explicit PAC-Bayesian bound proposed by [Debreko et al., 2004]. The idea is to bound the deviation between the two random variables $R_h(X_u)$ and $R_h(X_l)$, which are both concentrated around their mean $R_h(X_{l+u})$.

Let $p = p(X_{l+u})$ be a (prior) distribution over the class of binary hypotheses H that may depend on the full sample. Let $\delta \in (0;1)$ be given. Then, with probability at least $1-\delta$, the following PAC-Bayesian transductive bound holds for any $h \in H$,

$$R_h(X_u) \le R_h(X_l) + \sqrt{\left(\frac{2R_h(X_l)(l+u)}{u}\right)\frac{\log\frac{1}{p(h)} + \ln\frac{l}{\delta} + 7\log(l+u+1)}{m-1}} + \frac{2(\log\frac{1}{p(h)} + \ln\frac{l}{\delta} + 7\log(l+u+1))}{m-1} \qquad (7.\ 17)$$

In this formula, Debreko et al. [2004] demonstrated that this bound is located between the training error and the error over the full dataset. Also, we observe that the in the best possible scenario the transduction risk is equal to the training error. Further, Debreko et al. [2004] derive the following corollary:

Corollary 7. 1. Let A be any clustering algorithm and let h_τ, $\tau = 2,....,c$ be classifications of test set X_u as determined by clustering of the full sample X_{l+u} (into τ clusters). Let $\delta \in (0; 1)$ be given. Then with probability at least $1-\delta$, for all τ, (7.17) holds with $\log(1/p(h))$ replaced by τ and $\ln(m/\delta)$ replaced by $\ln(mc/\delta)$.

This extension is useful in situations, where the prior knowledge about ideal τ in unavailable. Further, Debreko et al. [2004] extended Corollary 7.1 to evaluate an *ensemble* of clustering algorithms. Specifically, we can concurrently apply k clustering algorithm (using each algorithm to cluster the data into $\tau=2,....,c$ clusters). We thus obtain kc hypotheses (partitions of X_{l+u}) and replace $\ln(cm/\delta)$ by $\ln(kcm/\delta)$ in Corollary 7.1 to guarantee that these bounds hold simultaneously for all kc hypotheses (with probability at least $1-\delta$). We thus choose the hypothesis, which minimizes the resulting bound. This extension is particularly attractive since typically without prior knowledge we do not know which clustering algorithm will be effective for the dataset at hand.

7. 6. 2 Model Selection

Clearly, the overall performance of the serial multi-expert framework relies on the performance of transductive inference implemented within the individual experts, which in turn depends on the quality of the generated clusters. There are two sets of parameters required to generate cluster model. First, these are the parameters required by clustering techniques. Second, it is cut-off thresholds for the feature discriminative scores. Parameters required by clustering techniques include distance metrics for the distance-based clustering techniques, number of mixture components for probabilistic clustering techniques etc. The cut-off theshold is required to select only highly scored features of the brushwork classes relevant to the expert sub-task based on the preprocessed matrix of feature scores. We discuss construction of this matrix in Section 7.5.2(b). To calculate a pool of clustering models we perform semi-supervised clustering using varying clustering parameters and cut-off thresholds. However, it is unclear which cut-off threshold and clustering parameters would be the most appropriate model for the particular sub-task. To choose such model, individual expert performs the model selection step as demonstrated in Figure 7.5.

Input:

1. A full sample set X_{l+u} and training sample set X_l,
2. Feature weighted scores $F_S(L_j)$ for the candidate class labels L_j,
3. A maximum number of mixture components or clusters K,
4. A set of cut-off thresholds for the feature weighted scores T_f

Output:

Candidate class labels of the test set X_u

Algorithm:

For each cut-off threshold $tf \in T_f$

For each number of mixture components or clusters k, $2 \leq k \leq K$

train cluster model $\{M_{k,tf}\}$ on X_{l+u} ;

For each model $\{M_{k,tf}\}$

Generate corresponding hypothesis $\{h_{k,tf}\}$ by estimation data clusters

(for each data point we perform maximization of the posterior probabilities with respect to the calculated distributions)

Measure cluster purity of each cluster

Calculate PAC-Bayesian bound of $\{h_{k,tf}\}$

Figure 7. 5. The model selection step

7. 7 Experiment Results

In this section we demonstrate the performance of the multi-expert framework for the brushwork annotation task. For our experiments, we randomly select 30 paintings from the subset of 315 paintings as discussed in Section 5.3. The selected paintings span such painting styles as: Renaissance, Fauvism, Impressionism, Post-Impressionism, Expressionism and Pointillism. From these paintings we extract 4880 fixed-size blocks of size 32x32. We further randomly split this dataset of image blocks; we employ 25% of the dataset for testing and 75% for training. Figure 7.6 demonstrates the distribution of brushwork classes in the training and testing sets.

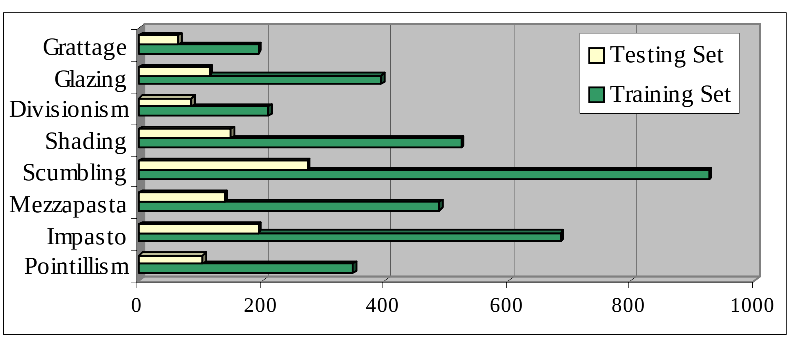


Figure 7. 6. Distribution of the brushwork class labels in the dataset

7. 7. 1 Automatic Feature Selection

To perform annotation of brushwork classes we combine low-level color and brushwork features discussed in Section 7.3 and meta-level artistic color features discussed in Chapter 6. We calculate feature scores using Chi-Square characteristics with respect to each individual feature. However, for demonstration purposes we organize the total feature set of 426 features in the feature groups as follows: Directional edge histogram (ED), edge pixels (EP), directional tilt histogram (TH), Gabor-based features (G), wavelet-based features (W), multi-scale fractal dimension (MFD), fractal dimension and lacunarity in HSI color space (FDL), major colors with account for perceptual similarity (MCP), color contrasts (CST), color histogram statistics (CHS), statistics of image magnitude (M) and Zernike moments (Z). For each group we calculate its average feature discriminative score using its respective features. Figure 7.7 demonstrates the plot of the averaged feature scores for each group with respect to the brushwork classes.

From Figure 7.7 we can observe that features in Edge Histogram and Edge Pixel group have the highest importance for such classes as *impasto, grattage* and *shading*. This is not a surprising result since patterns of *impasto, grattage* and *shading* exhibit a large number of linear components. Tilt Histogram features capture the properties of image gradients in terms of their strength and orientation. These features score highly with respect to the classes of *mezzapasta, divisionism* and *impasto*. Such scores are in line with relationships among classes and features as presented in Table 7.1, since *mezzapasta* and *divisionism* exhibit nearly no direction at all in contrast to *impressionism*.

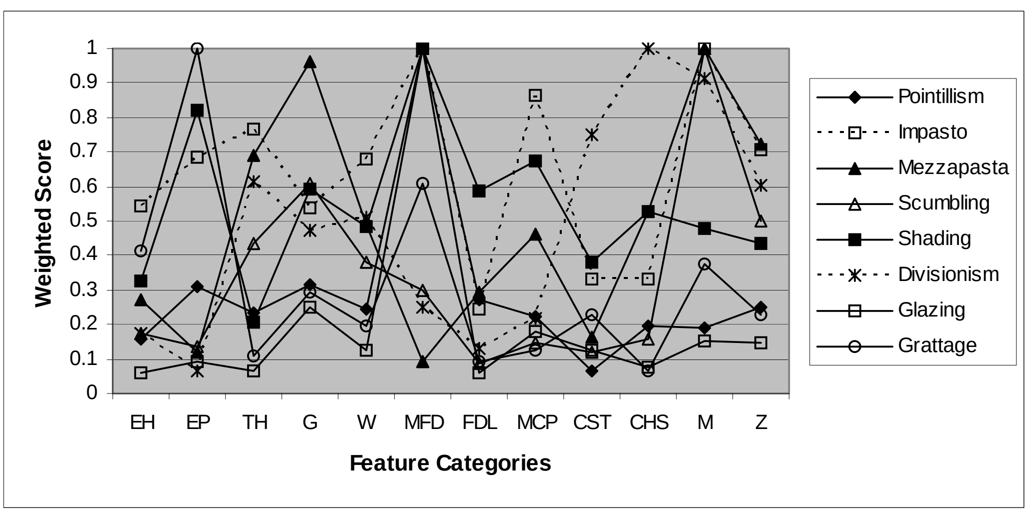


Figure 7. 7. Averaged feature scores of feature groups

At the same time, *mezzapsata* has gradients of low strength in contrast to *divisionism*. Gabor features have high importance for *mezzapasta* class. Figure 7.7 shows that the wavelet-based features do not exhibit very high importance for any of the classes. This is due to the averaging of scores within this group, where only a small fraction of features has high importance as compared to the rest of features. Wavelet-based features have relatively high importance for *divisionism, scumbling, shading* and *mezzapasta* classes. Multi-scale fractal dimension features score highly with respect to *shading, glazing, pointillism* and *impasto* due to the fact that these features represent well non-homogeneous patterns such as *shading* and *glazing* and patterns with high degree of roughness such as *pointillism* and *impasto*. Major colors with account for perceptual information (MCP group) are relatively more important for *impasto* and *shading* classes, since artists often used these brushwork techniques to depict sky, grass as well as dark-colored folds in medieval paintings. Color contrast and color information is important for *divisionism* since it exhibits a mixture of contrasting colors (or color mixing principle). Statistics of image gradient magnitude is naturally important to brushwork classes exhibiting distinctively high or low gradient magnitudes such as *impasto, divisionism* and *mezzapasta*.

7. 7. 2 Annotation Experiments

In our experiments, we test the proposed approach using several configurations of multi-level serial framework and compare it with several baseline methods. This includes:

- Baseline system (BS) is a one-step semi-supervised clustering method that utilizes full feature space. In our experiments we found that the use of 50 clusters and 30 mixture components yields the best results for distance-based and probabilistic clustering methods respectively. Therefore, we initialize the clustering techniques for baseline system with these values.
- Baseline system with automatic feature selection (BAFS) is similar to BS but it utilizes a reduced set of relevant features selected based on the feature scores as discussed in Section 7.5.2(b);
- Multi-expert framework with model selection (MMS) denotes the proposed transductive inference framework with model selection step as discussed in Section 7.6.3;
- Multi-expert framework with manual feature selection (MMFS) denotes the proposed transductive inference framework that utilizes model selection step and instead of automatic feature selection utilizes manual feature selection as discussed in Section 7.5.2(a);

We perform the testing of each configuration of the multi-expert framework with respect to Class Reduction and Class Reevaluation Strategies. For each of the developed configurations, we test performance based on the several clustering techniques implemented within individual experts. These techniques include: K-means clustering, ‘Complete-link’, ‘Average-link’, ‘Single-link’ agglomerative clustering and probabilistic clustering using a combination of Gaussian Mixture Model and Expectation Maximization.

System	Configuration	K-means	Complete	Average	Single	**GMM+EM**
Baseline	BS	74.61%	74.73%	75.08%	57.64%	**80%**
	BAFS	78.39%	79.03%	79.10%	58.11%	**83.6%**
Multi-Expert Framework	MMFS, Class Reevaluation	87.2%	87.5%	88.15%	62.17%	**89.3%**
	MMFS, Class Reduction	91.4%	92.06%	94.89%	68.32%	**95.38%**
	MMS, Class Reevaluation	85.67%	85.45%	86.72%	60.87%	**87.45%**
	MMS, Class Reduction	**90.23%**	**89.57%**	**92.71%**	**65.13%**	**93.71%**

Table 7. 2. Annotation performance of brushwork concepts

Table 7.2 demonstrates performance of the baseline and multi-expert framework in terms of the overall annotation accuracy. It shows that both the baseline and multiple expert framework obtain significantly higher performance with Complete-Link, Average-Link and K-means distance-based clustering techniques as compared to the Single-Link technique.

Since the Single-Link method merges two clusters with the smallest minimum pair-wise distance, it tends to group together patterns of the different classes, leading to a large number of impure clusters. In many cases, Average-Link yields slightly better results as compared to Complete-Link and iterative K-means clustering. Probablistic clustering technique results in better accuracy as compared to the distance-based clustering techniques. Such improvement of accuracy is due to the use of more sophisticated distance metrics in probablistic clustering.

Table 7.2 shows that the multi-expert system achieves significantly better performance as compared to the baseline system due to several reasons. First, the multi-expert system facilitates step-wise disambiguation of the patterns using domain knowledge and, thus, minimizes the probability of misclassifications at the terminal nodes.

Second, the model selection step facilitates adaptive selection of the best performing model and contributes to improvement in the overall accuracy.

To discuss the first point in more detail, we plot Figure 7.8. This figure demonstrates how terminal nodes benefit from the disambiguation process. Here, the task of the expert associated with the current terminal node is to assign the input patterns to one of the two classes (*divisionism* or *impasto*). The *Input Set* in Figure 7.8 is the set of the unlabelled patterns given as the input to the current terminal node. It represents the more coarse decision of the ancestor node as compared to the current terminal node. From the point of view of the current expert associated with the terminal node, these unlabelled patterns are likely to be *divisionism* or *impasto*.

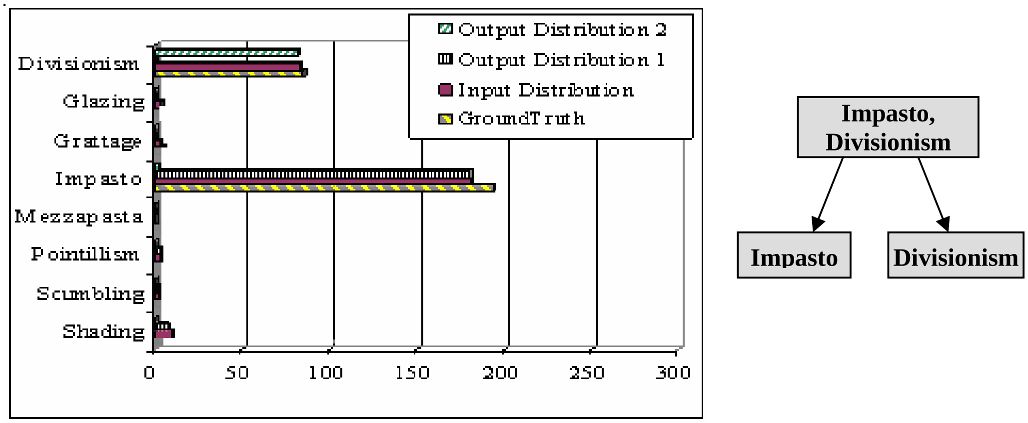


Figure 7. 8. Example of the terminal node

Figure 7.8 demonstrates the distribution of the unlabelled patterns in the *Input Set* with respect to their true labels (*Y* axis). It can be seen that the candidate class labels (*divisionism* or *impasto*) include the true class label for the majority of unlabelled patterns. Based on the input patterns, current expert generates its own decision and outputs the *Output Distribution 1* (*impasto*) and *Output Distribution 2* (*divisionism*). Figure 7.8 demonstrates that for the majority of unlabelled patterns their assigned labels *Output Distribution 1* and *Output Distribution 2* are in agreeament with their respective true labels *Ground Truth*.

From the distribution of unlabelled patterns in the *Input Set*, it is clear that the sequential refinement disambiguates patterns before they reach the current expert (terminal node) and receive their final label. This refinement naturally leads to higher accuracy achieved by the individual experts since the probability of the true class being assigned to the disambiguated patterns is high, resulting in better performance of the multi-expert framework.

Next, we discuss our second point in more detail. Model selection step performs selection of the least erroneous model and most appropriate cut-off threshold for individual sub-tasks, thus maximizing overall accuracy. In our experiment, we use several configurations to demonstrate that the use of relevant features and model selection enhances the annotation results. Initially, we combine baseline with automatic feature selection (BFS) discussed in Section 7.5.2(b). We used fixed cut-off thresholds for the feature scores at 0.7 level and found that the use of relevant features indeed improved the annotation result. This improvement is due to the dimensionality and noise reduction in the feature space. Next, we test multi-expert framework using model selection step with manual (MMFS) and automatic feature selection (MMS). Both configurations outperform baseline with automatic feature selection, while the performance of the multi-expert framework with model selection based on manual feature selection outperforms the same setup with automatic feature selection. However, their performance is comparable with around 1-3% loss of accuracy in MMS configuration.

Table 7.2 demonstrates that in all cases the use of Class Reevaluation strategy yields worse performance than that of the same setup with Class Reduction strategy. This is because under the Class Reduction strategy, some patterns receive their final labels at the intermediate nodes. Such conditional assignments result from high confidence of the experts at these nodes. However, the decision process annotates such patterns at the level of coarse intermediate decisions and disambiguation of these patterns is only partial, which results in additional 5% to 6% erroneous labels under Class Reevaluation strategy as compared to the Class Reduction strategy.

To conclude our experiments with brushwork annotation, we examine the distribution of annotation error of the multi-level framework with respect to the brushwork classes. Figure 7.9 plots the error rates of annotation based on the Class Reduction and Class Reevaluation strategies using Complete-Link, Average-Link and probabilistic clustering techniques with model selection. Figure 7.9 demonstrates that for all clustering techniques, Class Reduction strategy yields fewer errors in annotation for the majority of classes as compared to Class Reevaluation strategy. Also, we can observe that the distribution of error is non-uniform for all graphs. The majority of erroneous assignments are in *shading* and *scumbling* classes due to the fact that patterns in these classes exhibit a large variety of patterns and, thus, resemble other classes to a higher extent.

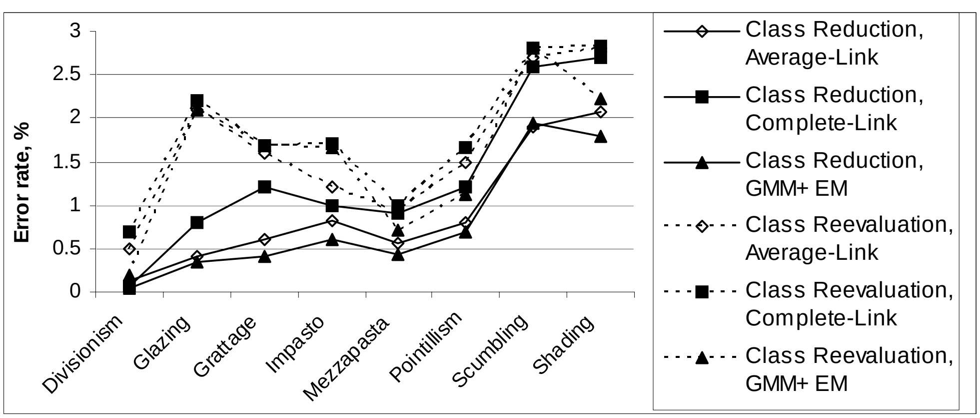


Figure 7. 9. Error distribution with respect to the brushwork classes

Both Class Reduction and Class Reevaluation strategy produce relatively smaller error for such classes as *divisionism, mezzapasta* and *pointillism*. This is due to the fact that patterns of these classes exhibit less variety and are adequately represented by a number of texture features, resulting in low intra-cluster pair-wise distances in the feature space.

7. 8 Summary

In this section we proposed the semi-supervised multi-expert approach for the annotation of brushwork in paintings. Explicit annotation of brushwork is desirable since it helps in the annotation of paintings with higher-level semantic concepts such as the artist names, periods of art and paintings styles. To perform annotation, we employed serial combination of multi-experts. This framework benefits from sequential refinement of the assigned labels and it facilitates dimensionality reduction. It generates decisions in accordance to the decision hierarchy that predefines similarity among classes based on the domain knowledge. To facilitate annotation at the level of individual experts, we employed semi-supervised distance-based and probabilistic clustering techniques. These techniques model the brushwork classes as tight clusters in the feature space as well as benefit from the distribution of unlabelled patterns. We presented several versions of the proposed of framework, where the relevant features and model parameters are selected manually or automatically using the iterative model selection step. Experiment results demonstrate satisfactory performance of the several versions of the proposed framework. The framework version with automated selection of features and clustering parameters yields comparable results to the version with manually pre-defined relevant features.

Chapter 8

Annotation of Application-Level Concepts

8. 1 Introduction

In this chapter, we focus on the annotation of paintings with high-level semantic concepts. Annotation of general domain images with high-level concepts is an active research area in recent years with several studies. Most studies focused on the annotation of art images, including existing models for general domain image annotation [Barnard et al. 2003, Li et al., 2004]. These models perform mapping from low-level features directly onto high-level concepts. Similarly to these works, our approach employs the visual content to perform annotation of paintings. However, unlike many traditional annotation frameworks, it does not limit itself to the analysis of visual content. We perform the annotation of high-level concepts in two steps. First, we combine visual-level concepts and low-level features to annotate image blocks with high-level concepts. Second, we disambiguate the annotated concepts at the image level with an account of the ontological relationships among concepts.

To perform annotation of image blocks with high-level concepts we employ multi-expert transductive inference framework as discussed in Chapter 7. This framework employs serial multi-experts approach to perform the annotation of patterns. One of the key features of this approach is its ability to introduce domain-specific knowledge, which reflects the similarity among the concepts, into the annotation process. Within the serial multi-experts approach, such knowledge is depicted in the decision hierarchy, which guides the pattern disambiguation process.

In the previous Chapter, we proposed two variations of the multi-expert transductive inference framework that employ similarity-based and probabilistic clustering methods. We have discussed Gaussian Mixture Model along with its advantages in Section 7.6.3. To perform annotation of high-level concepts we employ this probabilistic clustering method because the “soft” clustering result achieved via this model facilitates the detection of outliers.

To perform concept disambiguation, we utilize ontological relationships among concepts as discussed in Chapter 4. This ontology includes two types of concept relationships. First type is "parent-child" relationships between the meta-level and application-level concepts, which we exploit for the generation of the semantic labels. The second type is the relationships between high-level concepts as discussed in Section 4.5. We view this type of relationships among high-level concepts as constraints. To satisfy these constraints, we employ the integer linear programming (ILP) apporach. This method is somewhat similar to the one used for the semantic role labeling [Punyakanok et al., 2004; Tsai et al., 2005] task from the text-processing domain.

8. 2 Related Work

In this chapter, we perform automatic annotation of high-level concepts using domain ontology. In this respect, this approach is similar to the works of [Fan et al., 2005a and 2005b; Mylonas et al., 2006; Petridis et al., 2006; Dong et al., 2006]. These studies employ machine learning techniques to perform annotation of images with ontology concepts. Mylonas et al. [2006], Grira et al. [2005] employed agglomerative clustering to perform semi-supervised inference of image labels, while Li and Sun [2006] utilized 2D Conditional Random Fields for this purpose. A number of authors, including Zhao et al. [2005], Miller et al. [1997], Nigam and Mccallum [2000], Fang et al. [2005] and Miller et al. [2003], employed mixture models for the concept annotation task using both labeled and unlabeled data.

In our work, we aim to develop robust classifiers since the account of outliers is crucial. From this point of view, our work is related to the studies of Dave et al. [1991] and Miller et al. [2003]. In their study, Dave et al. [1991] introduced a "noise" cluster to capture outliers that aims to reduce contamination of true clusters. Our approach is somewhat more similar to the work of Miller et al. [2003], where multiple noise clusters are allowed in the semi-supervised setting. Nigam et al [2000] proposed another solution to handle outliers in unlabeled datasets. They gave a different (constant) weight to unlabelled instances in an attempt to reduce the influence of outliers on the annotation accuracy. However, this method treats all unlabeled instances in the same way and, thus, diminishes their impact on the estimation. Tajudin et al [2000] proposed an improvement by adopting a mixture modeling approach, where variable weights are given to each unlabeled sample. In contrast to the above discussed methods, we do not model the outliers explicitly, but rather implicitly re-evaluate them within our multi-expert framework. Rahman et al. [1999] and [2000] discussed a topology of multi-expert approaches. Closely related to the multi-expert approach are the decision combinations methods discussed in Ho et al. [1994].

In our work we employ domain knowledge in the form of ontology-provided constraints to improve the automatically generated labels. This is a relatively new area of research in image auto-annotation task. The majority of image studies employ constraints embedded in the training set [Zhou et al., 2005] or utilize user-provided pair-wise constraints [Grira et al., 2005] to improve the annotation accuracy. In our work we pose this problem as an optimization task that aims to generate the annotations, which are both consistent with ontology-provided constraints and the confidence values generated by the auto-annotation framework. To our knowledge, this approach has not been used for ontology-based annotation of images.

8. 3 Annotation of Application-Level Concepts

This section discusses the last stage of the framework proposed in Chapter 5. To perform the annotation we: (a) auto-annotate images based on the calculated features, and (b) utilize domain-specific knowledge to disambiguate automatically the generated results. We discuss these two stages in section 8.3.1 and 8.3.2 respectively.

8. 3. 1 Transductive Inference of Application-level Concepts

To auto-annotate images with high-level concepts, we perform a three-step procedure. First, we sub-divide paintings into fixed size blocks and perform iterative K-means clustering of painting blocks using low-level color and texture features. This procedure merges the dataset into clusters and represents similar image blocks as a single discrete data point, thus reducing the computational time. This is especially important for transductive inference methods: it might require long time to build a model using thousands of unlabelled samples. We represent the calculated feature clusters using mean feature vectors found within each cluster and utilize the feature clusters as units of analysis in the annotation task.

Second, we perform annotation of visual-level concepts. For the annotation of a cluster with brushwork concepts, we utilize low-level color and texture features of a cluster and employ a fully automated variant of the multi-expert transductive inference framework proposed in Chapter 7. To annotate the visual-level color concepts, we employ the methods discussed in Chapter 6. We first perform annotations of color temperature, color palette and contrasts for fixed size image blocks. To capture details of the color distribution, we measure color contrast within each block. For this, we form color pairs based on the major colors of a block and employ geometrical relationships among these colors to measure color contrasts. Next, we utilize the majority vote strategy to assign color concepts to clusters.

Third, we combine low-level color and texture features, and annotated visual-level concepts to map the feature clusters onto the art period, painting style and artist name concepts. We perform the mapping using a variation of the multi-expert framework based on the Gaussian Mixture Models proposed in Chapter 7. While the aim of our method is the correct classification of feature clusters into a set of "known" classes, we also aim to detect outliers and filter out the samples that belong to several "known" classes simultaneously. This need arises from the data itself. First, the altered appearance of brushwork and color concepts along the painting canvas, the object edges etc. Forcing such data samples to be annotated with semantic concepts might lead to classification error. Second, a painting exhibits a combination of several meta-level concepts and, naturally, some data samples from various paintings are likely to be ambiguous. For example, the data samples extracted from the background of paintings usually represent flat brushwork with almost homogeneous color. Such data samples are, therefore, not representative of particular artist, painting style and art period. We assume that only a subset of blocks is informative about artist, painting style and art period of an image. The probabilistic soft clustering generated by GMM facilitates the detection of ambiguous and rejected samples based on the posterior probability and the cluster purity measure. The transductive inference framework re-evaluates such patterns within the decision hierarchy as discussed in Section 7.5. However, this approach does not guarantee to eliminate all errors. To study the performance of the transductive inference framework closely, we evaluate this framework based on varying subset of image blocks, where the subsets arise from thresholding of the posterior probability of blocks.

To adopt the transductive inference framework, we preprocess the class weighted feature scores and we pre-define the decision hierarchy for the concepts of art period, artist name and painting styles respectively. Further in this section, we discuss the decision hierarchies predefined for the annotation of application-level concepts based on the time-line of art as discussed in Section 4.5. Figure 8.1 demonstrates the decision hierarchy for artist name concepts and Figure 8.2 shows the decision hierarchy for the painting style concepts.

Since our collection includes paintings from only two periods of art, the decision hierarchy has only three nodes: a root node and two leaf nodes. Due to this, the multi-expert framework becomes a single expert that annotates the image clusters with one of the two mutually exclusive concepts

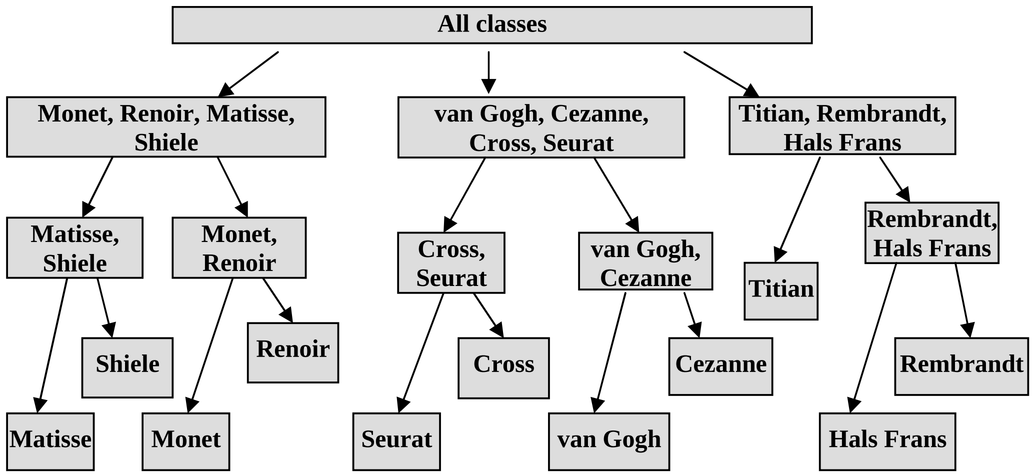


Figure 8. 1. The decision hierarchy for annotation with artist names

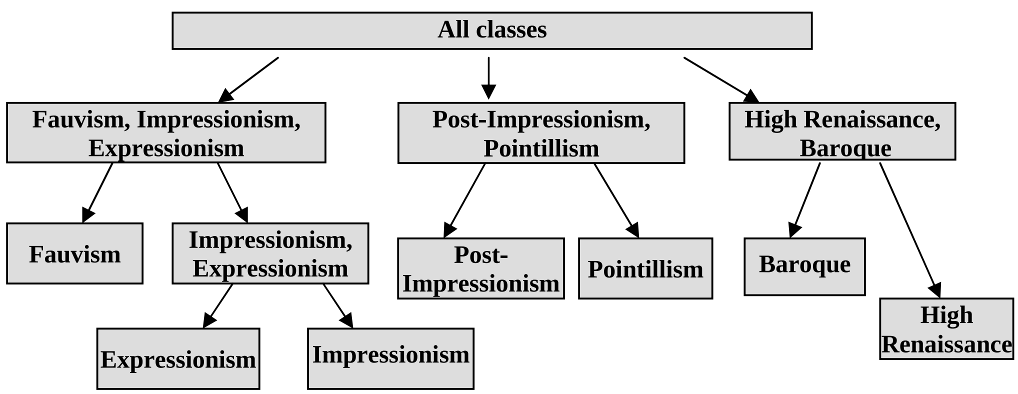


Figure 8. 2. The decision hierarchy for annotation with painting styles

8. 3. 2 Concept Disambiguation using Ontological Relationships

In this Section, we propose a method that integrates the generated high-level concepts and disambiguates them at the image level. While it is difficult to incorporate concept relationships during the learning phase, it is possible to account for these relationships after base classifiers generate their candidate concept labels. Ideally, if the learned base classifiers are perfect, blocks will be labeled correctly according to the classifiers' predictions. In reality, labels assigned to blocks in an image often contradict each other, and violate the constraints arising from domain knowledge. In order to resolve these conflicts, we design a disambiguation method that takes the confidence scores of each individual concept given by the base classifiers as input, and outputs the best global assignment that also satisfies the domain knowledge constraints. In domain knowledge, ontological relationships among the application-level concepts serve as such constraints as demonstrated in Table 4.6. For example, van Gogh's paintings appeared in Modern art period, but not in Medieval art period. To perform global optimization of labels, we propose an ontology-based concept disambiguation method (OCD) that is similar to the more general problem of the metric labeling proposed by Kleinberg et al. [2002]. To solve this problem we encode the concept relationships as linear constraints and employ the Integer Linear Programming approach [Chekuri et al., 2001]. Integer Linear Programming is a class of constraint satisfaction problems, where variables are restricted to the integer representation form. The goal of such a problem is to minimize (maximize) the *n-ary* function f, which is defined as the sum of variables c_iX_i. Figure 8.3 demonstrates the high-level scheme of the proposed OCD method.

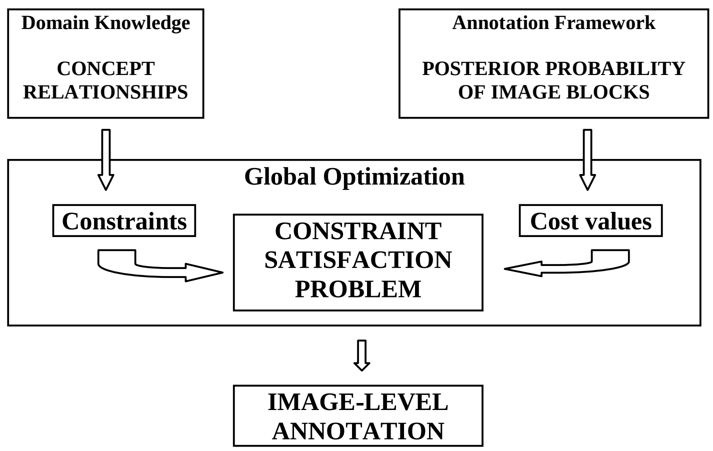


Figure 8. 3. Ontology concept-based disambiguation method

Overall, the proposed OCD method includes two stages. First, we calculate high-level concepts for the whole image and form label combinations. Second, using these combinations we solve a constraint satisfaction problem to generate the final labels for images.

To achieve the image-level representation of labels, we back-project the labels of blocks onto their respective images and calculate their distribution within an image. We represent such distributions using a histogram, where the histogram bins refer to the application-level concepts and the histogram values denote the number of image blocks annotated with these concepts. Using this method, we represent the distributions of artist name concepts, painting style concepts and art period concepts within each image. Based on these distributions we form all possible concept combinations <artist name – painting style – art period>.

Next, we aim to disambiguate these concepts and generate the final image-level annotations. Our disambiguation approach relies on the formulation of [Roth et al., 2004] and [Marciniak et al., 2005], who applied integer linear programming for the semantic role labeling problem in the Natural Language Processing domain.

In accord to the formulation of [Marciniak et al., 2005], the final decisions are modeled as a set of n classification tasks $CT = \{CT_1, \ldots CT_n\}$, which form inter-related pairs. We have three classification tasks: artist name, painting style and art period respectively. Each task CT_i assigns a label from the set $L_i = \{l_{i1},....l_{im_i}\}$ to an image. We model the assignments as the variables of linear cost function. We have simple variables that model assignments of each individual label and combined variables that model assignment of labels for each pair of related tasks. Thus, simple variable $x(l_{ij})$ models the individual assignments of every label in L_i for task CT_i. This label $x(l_{ij})$ is set to 1 if selected or zero otherwise. Each individual assignment $x(l_{ij})$ is associated with the assignment cost, which is defined as follows:

$$cost(l_{ij}) = -log_2(p(l_{ij}*w(l_{ij})) \quad (8.1)$$

where $p(l_{ij})$ denotes the mean posterior probability generated by the multi-expert framework system for image blocks that are labeled with concept l_{ij}; $w(l_{ij}))$ is the normalized number of such blocks within an image.

Combined variable $x(l_{ij},l_{kt})$ models the assignment of labels between two inter-related tasks CT_i and CT_k. This variable is equal to 1 if our method attempts to annotate an image with concepts l_{ij} and l_{kt} and 0 otherwise. Each of these assignments is associated with a coefficient that reflects the domain constraint on the respective pair of labels. The value for this variable arises from the acyclic graph H that we employ to represent the ontological relationships between artist name, painting style and art period. If two concepts l_{ij} *and* l_{kt} in H are related, then $H(l_{ij},l_{kt}) = 1$ and otherwise it is set to 0.0001. We calculate the coefficient in the following way:

$$coef(l_{ij},l_{kt}) = -log_2(H(l_{ij},l_{kt})) \quad (8.2)$$

The OCD method includes the target function and a set of constraints that prohibit illegal assignments. In our case, the target function is the cost function f, which we want to minimize:

$$\min f = (\sum_{CT_i \in CT} \sum_{l_{ij} \in L_i} \cos t(l_{ij}) \times x(l_{ij}) + \sum_{CT_i, CT_k, i<k} \sum_{l_{ij} \in L_i, l_{kt} \in L_k} coef(l_{ij}, l_{kt}) \times x(l_{ij}, l_{kt})) \quad (8.3)$$

We formulate several constraints. First, the algorithm should select exactly one label l_{ij} for each task CT_i. Thus only one variable x can be set to 1:

$$\sum_{l_{ij} \in L_i} x(l_{ij}) = 1, \quad \forall i \in \{1...n\} \quad (8.4)$$

We also require that if the two variables $x(l_{ij})$ and $x(l_{kt})$ are selected, then exactly one combined variable $x(l_{ij},l_{kt})$ that models the co-occurrence of these labels must be set to 1:

$$x(l_{ij}) - \sum_{l_{kp} \in L_k} x(l_{ij}, l_{kt}) = 0,$$
$$\forall i, k \in \{1,..,n\}, i < k \wedge j \in \{1,...,m_i\}$$

$$x(l_{kt}) - \sum_{l_{ij} \in L_i} x(l_{ij}, l_{kt}) = 0, \qquad (8.5)$$
$$\forall i, k \in \{1,..,n\}, i < k \wedge t \in \{1,...,m_k\}$$

Lastly, we pre-define that variables x and y are binary:

$$x(l_{ij}) \in \{0,1\} \wedge x(l_{ij}, l_{kt}) \in \{0,1\},$$
$$\forall i, k \in \{1,..,n\} \wedge j \in \{1,...,m_i\} \wedge t \in \{1,...,m_k\} \qquad (8.6)$$

8. 4 Experiment Results

For our experiments, we employ the full dataset of 1050 paintings as discussed in Section 5.3. We employ 315 and 735 images for training and testing respectively. For annotation of the application-level concepts we utilize 32x32 fixed-size blocks of size. To achieve the image clusters, we allow up to 60 clusters in each painting.

To present our experiment results, we plot all precision, recall and F1 values discussed in Section2.5 with respect to the increasing number of the rejected blocks. At each level of the rejection rate, we reduce the number of the analyzed image blocks in accordance to their confidence value, which is the generated posterior probability. For example, 10% rejection rate means that we discard 10% of the least confident samples and evaluate the performance based on the remaining 90% of the whole sample set. By varying the percentage of rejected blocks, we could demonstrate that the generated posterior probabilities are reliable, and evaluate the impact of using only a subset of most reliable blocks to induce high-level semantics at the image level.

8. 4. 1 Annotation of Artist Concepts

In this section, we evaluate the proposed framework with respect to the artist name concepts. First, we evaluate the annotations generated for the image blocks. Next, we evaluate the image-level annotations. Lastly, we evaluate the performance of the proposed framework with respect to each artist and investigate the dependencies between the size of the training dataset and the annotation accuracy.

Figure 8.4 demonstrates the precision of the block-level annotations generated by the proposed multi-experts framework and several baseline methods. To calculate precision, we compare the ground truth of image blocks and their candidate labels, which are generated in the leaf nodes of the decision hierarchy of the transductive inference framework. At each level of rejected rate, we calculate the number of correctly annotated blocks and normalize this number by the number of currently analyzed blocks. The baseline methods include: 1) Baseline 1 – inductive inference based on low-level features; 2) Baseline 2 – transductive inference based on low-level features; 3) Baseline 3 - inductive inference based on low-level features and visual-level concepts; 4) Baseline 4 - transductive inference based on low-level features and visual-level concepts. For inductive inference, we employ the multi-category probabilistic SVM method proposed by Chakrabartty et al. [2002]. For transductive inference, we employ the combination of GMM and EM using 150 distributions. For all baselines, we employ the 100 top-scoring features based on the Chi-square statistics.

Several observations can be readily obtained from Figure 8.4. First, it shows that the precision of all methods improves with the increasing rejection rate, since we increasingly remove ambiguous samples from the dataset and decrease the number of analyzed blocks. Second, it shows that visual-level concepts (Baseline 3, 4 and the proposed method) facilitate higher annotation accuracy as compared to the use of low-level features only. Third, it shows that for all Baselines the transductive inference method slightly outperforms inductive inference.

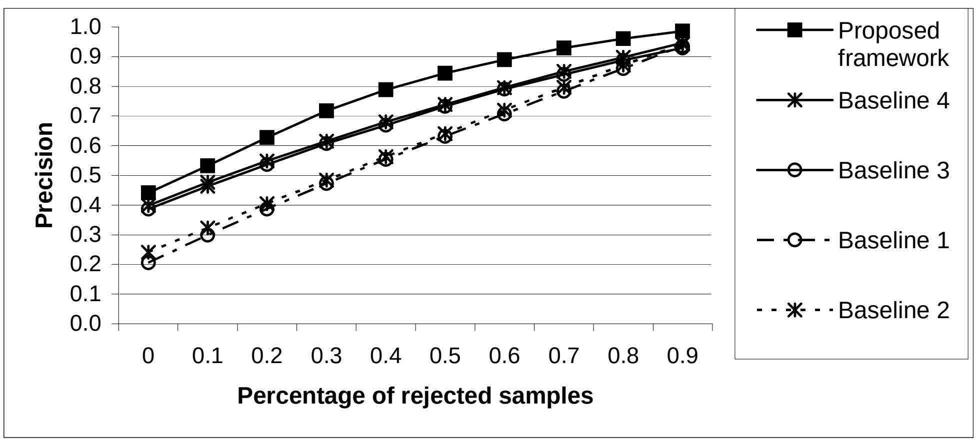


Figure 8. 4. Region-based annotation performance for artist name concepts

Fourth, the precision graph of the proposed method demonstrates superior results due to several factors. They include: (a) the use of visual-level concepts leads to the increased performance; (b) the decision hierarchy facilitates step-wise disambiguation of patterns, which purifies the classifier decision and improves the predicted accuracy of classification; and (c) the model selection step adaptively selects features and model parameters and finds the most adequate model to capture the data distribution.

Figure 8.4 demonstrates that the proposed method achieves accuracy of more than 90% for the rejection rate of 0.6 and higher. However, this graph accounts for the overall performance and does not emphasize the performance of individual categories. In Figure 8.5, we compare micro- and macro- precision. Micro precision reflects the performance over whole dataset, while macro precision demonstrates the performance accounts for the data distribution in individual classes.

From this Figure, we observe that the performance of individual categories is not uniform. We hypothesize that macro and micro precision graphs differ for several reasons. First, the training set for the artist name classes varies in size. Second, some classes are likely to have higher prediction accuracy due to the fact that they exhibit distinctive brushwork.

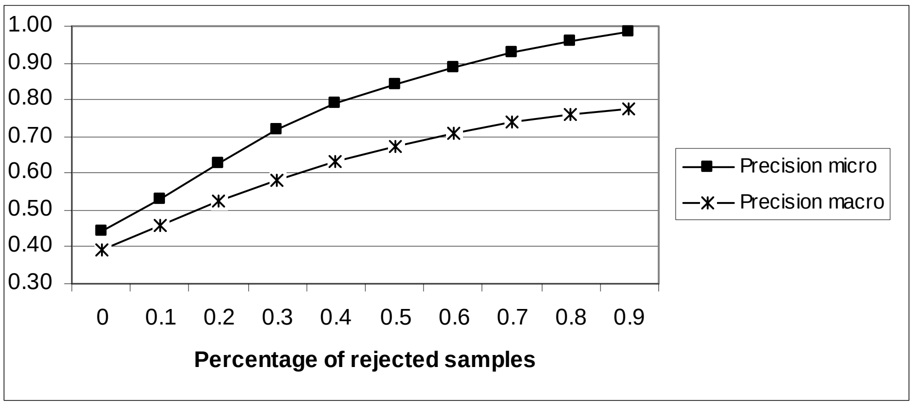


Figure 8. 5. Micro and macro precision of block-level annotation

For example, paintings by *Matisse* and *Schiele* exhibit a lot of *mezzapasta* brushwork technique. From Table 7.9 we observe that this technique has a relatively high recognition rate as compared to the other techniques. Since the brushwork technique is representative of the artist and painting style, the relatively high recognition rate of *mezzapasta* technique is likely to contribute to the recognition in *Matisse* and *Schiele* classes. Lastly, highly confident blocks are not distributed evenly across all classes. This difference becomes more apparent at high levels of the rejection rate, where several classes have most of their blocks rejected, thus resulting in zero level of precision and recall for those classes.

In Figure 8.6 we assess the image-level annotations using the majority vote strategy. In this graph we calculate micro recall, precision, F1 measure and macro F1 measure. Similarly to the previous figure, the axis *X* denotes the percentage of rejected blocks. To calculate recall and precision, we assume the correctly annotated images as *relevant*, annotated images are *retrieved*.

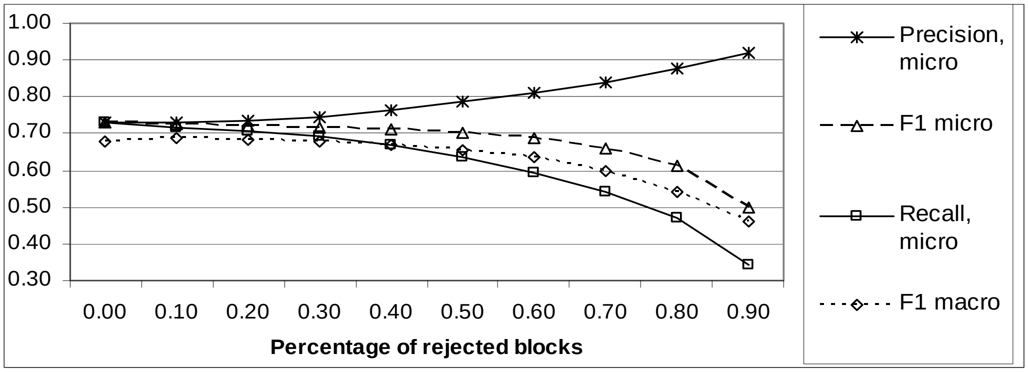


Figure 8. 6. Image-level annotation with artist name concepts

The curves for micro and macro F1 measures are close to each other, which means that the performance of both small and large artist classes is comparable. To better understand the performance of different classes, we tabulate in Table 8.1 the performance with respect to the individual categories using the F1 measure.

Rejection, %	Class										
	1	2	3	4	5	6	7	8	9	10	11
0.3	0.80	0.81	0.81	0.75	0.55	0.67	0.62	0.48	0.59	0.57	0.82
0.4	0.79	0.81	0.80	0.75	0.53	0.63	0.60	0.48	0.57	0.55	0.83
0.5	0.76	0.80	0.81	0.74	0.50	0.62	0.59	0.49	0.55	0.54	0.82
0.6	0.72	0.80	0.80	0.73	0.50	0.61	0.55	0.49	0.48	0.52	0.78

Table 8. 1. Performance in individual categories for artist name concepts

We demonstrate the performance at increasing levels of the rejection rate. It can be seen that classes 5, 8, 9 and 10 perform worse than other classes. These classes correspond to Titian, Rembrandt, Frans Hals and Cezanne. There are at least two reasons for this. First, some artists might have significantly higher variance of paintings like Cezanne. Second, some classes have small number of training samples like Titian, Rembrandt and Frans Hals. To investigate the relationship between the training size and performance, we generate Figure 8.7, which shows the distribution of the training sizes across all categories and the F1 measures achieved based on the full sample set, without any rejection.

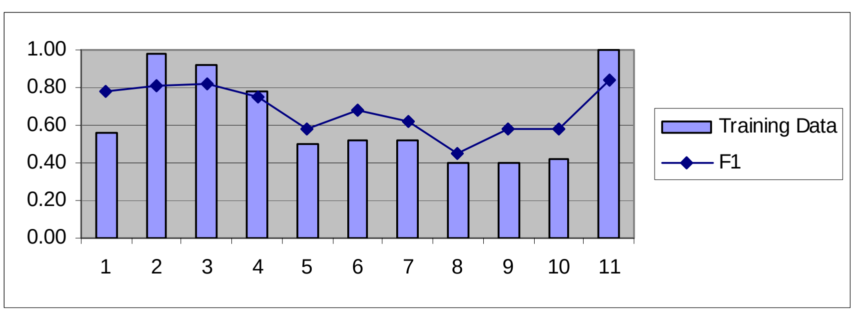


Figure 8. 7. Relationship between the training set size and F1 measure

We normalize the number of the training samples by the largest number of training samples across all categories. This Figure demonstrates that the size of training dataset and performance are quite well correlated. Lastly, we compare the majority vote (MV) strategy and the proposed method for ontology-based concept disambiguation (OCD) with respect to artist name concepts. With this experiment we aim to demonstrate that information about painting styles and period of art serves to increase the annotation accuracy of the artist name concepts. Figure 8.8 demonstrates recall and precision values of both methods with respect to the number of rejected patterns.

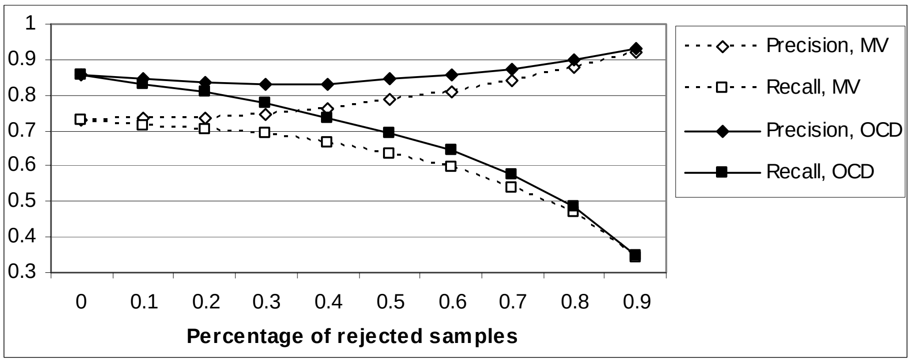


Figure 8. 8. Comparison of MV and OCD disambiguation for artist name concepts

From the Figure, we can observe that OCD method is most effective on the full dataset. This is due to the fact that each image has a large number of concept combinations based on the full dataset. With increasing rejection rate, the number of concepts combinations reduces, the majority vote strategy becomes less erroneous and both strategies deliver similar results for large number of rejected patterns.

While both strategies generate over 90% of precision for top 20% of image blocks, the use of OCD method is more beneficial for several reasons. First, it outperforms the majority vote strategy by around 10% for the full dataset and generates an accuracy of 85%. This result is comparable with the precision rate achieved by both systems for top 20% blocks. Second and most importantly, using OCD method for disambiguation based on full dataset, we are able to preserve the high recall rate. Third, since OCD method performs disambiguation within an image, it is not sensitive to the fact that the scale of confidence value across the semantic categories may vary. Lastly, this method combines the automated analysis of images and ontological relationships among concepts, which ensures that the system assigns the final labels in accordance to domain knowledge.

8. 4. 2 Annotation of Painting Style Concepts

In this section we evaluate the performance of the method with respect to the painting style concepts such as Impressionism, Baroque, Renaissance, Fauvism and others. In this experiment, we observed the same tendencies as in the annotation of artist name concepts, where macro-level statistics demonstrates that the performance of individual classes is slightly worse than the performance of the whole dataset. Figure 8.9 demonstrates the image-level performance of painting style annotation using the majority vote strategy.

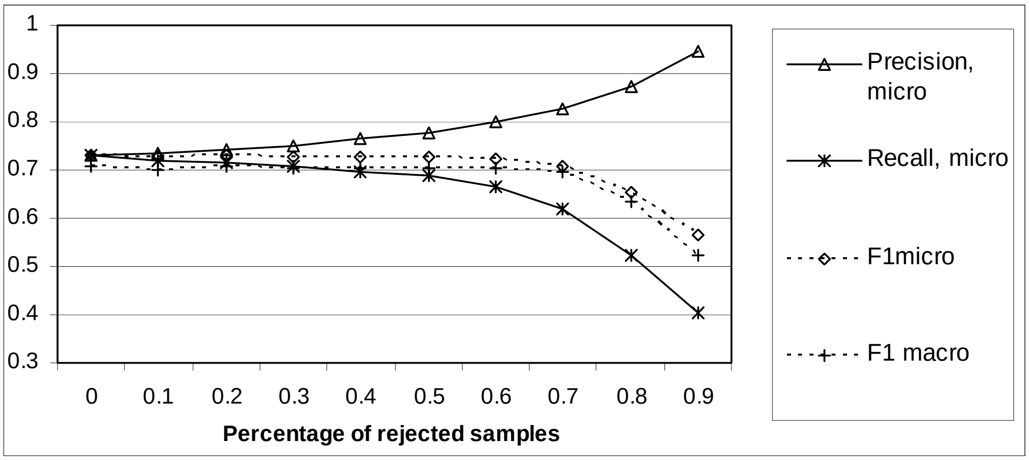


Figure 8. 9. Image-level annotation with painting style concepts

This graph shows the micro-precision, recall and F1 measures and macro-F1 measure. We construct this Figure similarly to Figure 8.4 with the axis *X* reflecting the percentage of the rejected samples. Evaluation of micro and macro F1 measures suggests that the categories with both large and small number of training samples perform comparably.

Table 8.2 again demonstrates the performance in individual categories using F1 measure. We show the performance in the individual categories based on the varying levels of the rejection rate. It can be seen that classes of Baroque and Renaissance shown in columns 6 and 7 of Table 8.2 perform worse then other classes. There are at least two reasons for this. First, these classes represent two styles of *Medieval* art and the use of brushwork and color information may not be sufficient to recognize these painting styles. Second, both classes have relatively small number of training samples in our dataset.

Rejection, %	Class						
	1	2	3	4	5	6	7
0.3	0.737	0.801	0.715	0.820	0.748	0.555	0.576
0.4	0.735	0.800	0.713	0.818	0.740	0.553	0.574
0.5	0.724	0.801	0.714	0.816	0.743	0.551	0.564
0.6	0.716	0.798	0.709	0.780	0.754	0.548	0.561

Table 8. 2. Performance in individual categories for painting style concepts

Figure 8.10 demonstrates the relationship between the number of training samples and the annotation performance. In this Figure, we normalize the number of training samples in the individual categories by the maximum number of training samples.

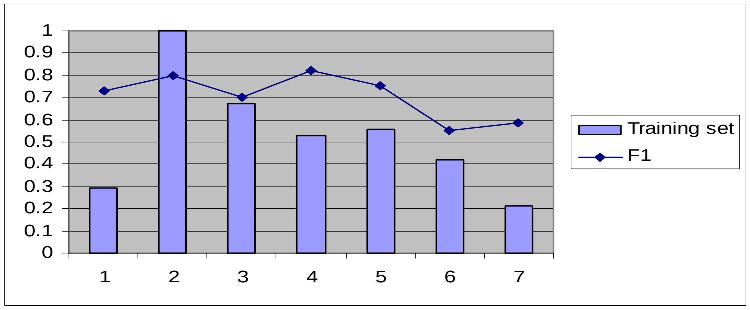


Figure 8. 10. Relationship between the training set size and F1 measure

We represent the annotation performance using F1 measure in each individual category achieved at the level of 0% of the rejected samples. Comparative performance of classes 2, 3, 4 and 5 with classes 6 and 7 demonstrates that relationship between the size of training dataset and annotation performance. However, somewhat surprising are the levels of F1 measure with respect to classes 1, 4 and 7. Close inspection of the dataset reveals that each of these classes includes a single artist. We hypothesize that these classes achieve relatively high F1 measure value due to the low variance of their respective images.

In Figure 8.11, we evaluate the performance of the majority vote (MV) and OCD disambiguation strategies for the annotation of painting style concepts to images. We calculate precision and recall values based on these strategies and plot these values with respect to the increasing number of rejected samples. We observe similar tendency as in Figure 8.6: OCD method has the highest relative performance with no rejection rate. This is due to the fact that full dataset has the largest number of concept combinations that makes disambiguation most effective.

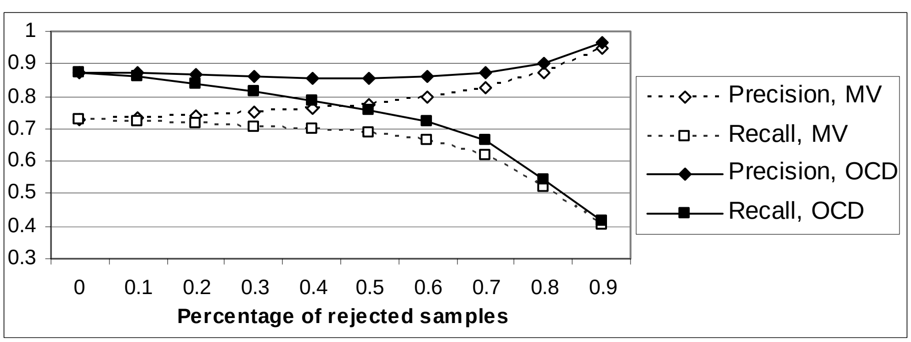


Figure 8. 11. Comparison of MV and OCD strategies for painting style annotation

Similar to the disambiguation of artist name concepts in Figure 8.7, this graph demonstrates that OCD strategy outperforms MV strategy on the full dataset. The use of OCD method is beneficial, since it generates around 87% of both precision and recall based on the full dataset.

8. 4. 3 Annotation of Art Period Concepts

We evaluate the annotation of art period concepts using several baseline systems. The first baseline system (Baseline 1) for our experiments is a multi-category SVM classification method based on low-level color and texture features. To test the contribution of the visual-level concepts to the overall accuracy, we employ the variation of the baseline system (Baseline 2) that combines meta-level artistic concepts and low-level features with class weighted feature scores above 0.7. Lastly, we evaluate the multi-expert transductive inference framework based on both low-level features and visual-level concepts. Table 8.3 demonstrates the performance of the systems. From these results, we draw the following observations. Baseline 2 results in higher accuracy as compared to Baseline 1 system due to several reasons. First, the use of visual-level concepts facilitates more accurate annotation as compared to the use of low-level features only. Second, the use of the weighted feature scores facilitates the reduction of the noise in the feature space.

System	Precision, % Block-level annotation	Precision, % Image-level annotation
Baseline 1	**68.72**	**81.48**
Baseline 2	**79.02**	**93.56**
Proposed framework	**86.84**	**98.71**

Table 8. 3. Annotation performance of art period concepts

Next, our proposed method achieves even higher accuracy of 98% at the image-level as compared to Baseline 2. The improvements arise from the use visual-level concepts, semi-supervised inference and model selection. Figure 8.12 illustrates the misclassified paintings. All of them belong to Modern art period. However, they all exhibit dark and red colors with large areas of mezzapasta brushwork class similarly to the paintings of Medieval art period.

The OCD disambiguation strategy reaches about 99.7% for the art period concepts.

Figure 8. 12. Examples of misclassifications for art period concepts

8. 4. 4 Ontology-based Concept Disambiguation

In this section we evaluate the proposed method for the concept disambiguation based on the ontological relationships. In this task, we are concerned with the correct annotations of artist name, painting style and art period concepts for each image. We consider an image as *retrieved*, if this image has automatically assigned labels of artist name, painting style and art period annotations for this image. If it predicts these labels correctly, we consider this annotation as *correct* or *relevant and retrieved*. In Figure 8.13 we plot the recall and precision values for images at increasing levels of the rejected rate. To plot the values, we perform rejection with respect to each individual category. Thus, at each step we reject X least confident patterns in each category rather than over the whole dataset. This will ensure that each category will retain patterns and the precision/recall metrics will have non-zero values in each individual category at high levels of rejection.

We compare two strategies: 1) majority vote (MV) and 2) the proposed method for the ontology-based concept disambiguation (OCD) as described in Section 8.3.2. This figure demonstrates that both recall and precision are higher for the proposed OCD method as compared to the majority vote.

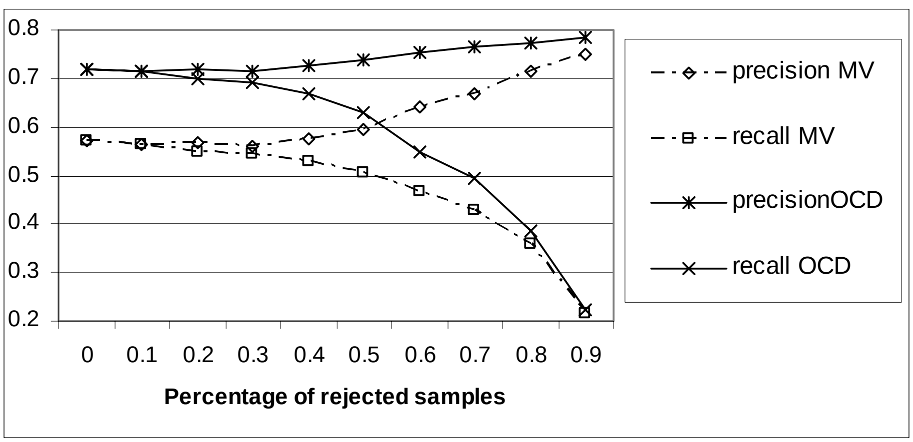


Figure 8. 13. Comparison of MV and OCD disambiguation methods

It further can be observed that with the increase of the rejection rate, the recall decreases for both strategies sharply. This is due to the fact that the rejection of image blocks leads to the increase of images that do not have all three concepts annotated. In turn, this results in the decrease of correctly annotated images.

The OCD method can be further improved in several ways. Consider the set of retrieved images. It can be formalized as *Retrieved = AdC+AdNC+NAd.* Here, *AdC* denotes the images that include at least one *admissible* concept combination and one of these combinations is correct in accordance to the ground truth. We define a combination as *admissible* if its concepts are related in accordance to the graph *H*. For example, the combination of "Da Vinci, Renaissance, Medieval" is admissible, while the combination of "Da Vinci, Abstractionism, Medieval" is not. The system should ideally annotate correctly all images that include the correct concept combination. *AdNC* denotes the images with admissible combinations but without the correct one. Lastly, *NAd* denotes the images with non-admissible concept combinations. Clearly, subsets *AdNC* and *NAd* are guaranteed to result in erroneous annotations for both MV and OCD methods, since our framework does not have additional information to disambiguate the concepts.

Using ontology-based concept relationships, the system can easily detect the subset *NAd* and attempt to re-evaluate annotations in this set. Alternatively, we can employ incomplete annotations that often accompany online painting collections. Incomplete annotations might contribute to the system performance in at least two ways. First, the use of incomplete annotations helps to reduce the set of admissible combinations derived from graph *H*. Second, incomplete annotations may serve to label *NAd* set of images. In our future work, we aim to focus on applying such incomplete annotations to improve the proposed OCD method.

In the rest of this section, we perform a preliminary experiment to evaluate a combination of OCD method with incomplete annotations. We assume that our collection has artist name annotations, for example, from the World Wide Web. The incorporation of Web annotations does not require modifications of the OCD method. To combine OCD method with Web data, we perform a two-step procedure. First, we generate the list of concept combinations based on the automatically generated block-level concepts. Second, we substitute artist name labels with the labels extracted from the Web data.

In Figure 8.14, we demonstrate the recall and precision of MV and OCD methods, and OCD method with Web data. It demonstrates that the use of incomplete annotations with OCD method results in the higher recall and precision rates as compared to both OCD and MV methods. This improvement is due to the fact that incomplete annotations offer additional disambiguation of the high-level concepts via the reduction of the admissible combinations.

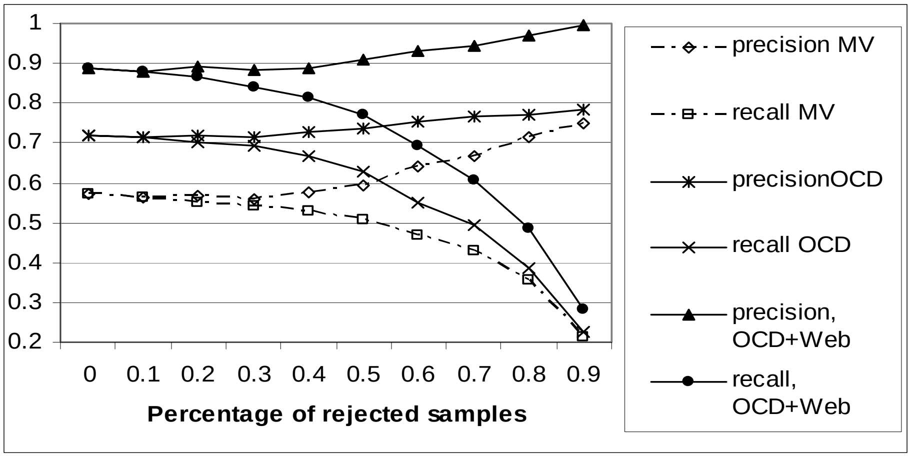


Figure 8. 14. Comparison of disambiguation strategies

Table 8.4 demonstrates the requirements of computational time for the proposed framework based on the datasets as discussed in the experimental setup.

Task	Time required
Extraction of features for the color analysis	6 hours
Annotation of Color Semantics	1.5 hours
Extraction of features for the annotation of brushwork and high-level semantics	48 hours
Annotation of Brushwork Classes	2.5 hours
Annotation of Artist Name	4 hours
Annotation of Painting Styles	3.5-4 hours
Annotation of Art Periods	0.5 hours

Table 8. 4. Computational time requirements

8. 5 Summary

In this chapter, we focused on the last stage of the proposed painting annotation framework. We proposed a method for annotation of paintings with high-level concepts that includes two major steps. First is the automatic generation of the candidate concepts. Second is the ontology-based concept disambiguation of the image-level labels. For the automatic generation of high-level concepts we employed the transductive multi-expert framework, which utilizes the domain knowledge in two ways. First, it combines the meta-level visual concepts and low-level features to annotate high-level concepts. Second, it generates the expert decision based on the decision hierarchy that encodes the concept similarity.

In our experiments, we showed that the multi-expert framework facilitates superior performance for the high-level concept annotation task due to the several reasons. First, the use of visual-level artistic concepts contributes to the annotation accuracy. Second, step-wise disambiguation and adaptive selection of features and model parameters facilitates higher recognition rates as we have demonstrated in Chapter 7.

Further, we proposed the concept disambiguation method that relies on the ontological relationships among concepts. We demonstrated that this method results in higher accuracy as compared to the widely used majority vote strategy. The proposed method easily extends to large number of related high-level concepts and facilitates the incorporation of incomplete annotations available from additional sources.

Chapter 9

Conclusions and Future Work

9. 1 Main Contributions

In this dissertation, we tackled the problem of automatic paintings annotation and retrieval using domain ontology. This dissertation spans several major theoretical areas of research and their application to multimedia: expert systems and semi-supervised learning. Expert systems represent a general approach in which a fully automated system performs concept inference based on the domain knowledge provided by human experts. Semi-supervised statistical learning performs automatic inference using both labeled and unlabeled data. The focus of this approach is to infer concepts based on the limited labeled set and compensate for the scarcity of the labeled set using a large number of unlabeled patterns. In this thesis we made four major contributions:

1. We introduced the framework for paintings annotation that combines the expert systems approach with supervised and semi-supervised statistical learning. This framework employs domain ontology and exploits various disambiguation strategies based on this ontology.
2. We proposed and implemented a method for annotation of paintings with artistic color concepts, which combines artistic color theory and inductive statistical learning techniques to annotate various color concepts.
3. We proposed and implemented the transductive multi-expert framework that performs step-wise disambiguation of concepts. The use of semi-supervised inference methods within this framework reduces the required number of labeled samples for effective learning.
4. We proposed and implemented the concept disambiguation method, which utilizes ontological relationships among concepts. We pose the problem of ontology-based

disambiguation as an optimization task, which is easily extendable to collections with incomplete annotations, for example, from WWW.

Further in this section we focus on the main contributions.

9. 1. 1 Framework for Ontology-based Annotation and Retrieval of Paintings

We proposed and implemented the paintings annotation and retrieval framework that relies on the domain knowledge. This framework exploits the three-level ontology of artistic concepts. Concepts of visual-level serve as the meta-level information that provides cues to the annotation of high-level concepts. High-level concepts aim to fulfill the user needs of both expert and novice users. We organized these concepts into abstract level and application level depending on the user needs. Our framework includes several stages of concepts annotation. First, it utilizes statistical learning to annotate visual-level concepts. Next, to annotate abstract-level concepts it performs ontology-based concept propagation based on the visual-level concepts. Lastly, it annotates the application-level concepts as a two-step process. First, it combines both low-level features and visual-level concepts to label image blocks with application-level concepts. Next, it integrates the block-level candidate labels to represent the whole image and utilizes the ontological concept relationships to disambiguate these labels.

In our experiments we demonstrated that the use of domain knowledge improves the annotation accuracy at various stages of the proposed framework. We demonstrated that the use of meta-level concepts within the proposed framework yields accuracy improvement of 10%-18% as compared to the same setup with low-level features only. Next, we demonstrated that the use of ontological relationships contributes to the more accurate concept disambiguation. The ontology-based disambiguation leads to the growth of both recall and precision by around 15% as compared to the majority vote method.

9. 1. 2 Method for Annotation of Artistic Color Concepts

We proposed a method for the annotation of paintings with artistic color concepts. This method relies on the artistic color theory that defines semantics of colors based on their geometrical and morphological relationships within color sphere. Our method performs region-based analysis of paintings, where each region is represented as multiple colors. The accounting for multiple colors facilitates the analysis of paintings from various paintings styles and periods of art and represents improvement over existing painting annotation studies. Another improvement is the set of domain-dependent features, which model distribution of various color temperatures and color palettes within a region. Our method then learns to relate various distributions of artistic color concepts within a region to the judgments about the whole image. The method further propagates visual color concepts via ontological relationships to achieve abstract-level annotations for the whole image. To evaluate performance of the proposed method, we implemented a paintings retrieval system using mid-sized collection of images downloaded from the World Wide Web. This system facilitates retrieval by both visual-level and abstract-level concepts. Our results demonstrate that the system yields satisfactory performance to a wide variety of expert-provided queries.

9. 1. 3 Semi-supervised Multi-Expert Framework

We proposed and implemented a framework that facilitates annotation of patterns using a multiple expert approach. The proposed framework organizes the whole task as several sub-tasks and encodes relationships among them within a decision hierarchy. Each node of the hierarchy is associated with individual experts that generate decisions using semi-supervised learning techniques. The advantage of this framework is that it performs step-wise disambiguation of patterns that might lead to improved accuracy. Further, this framework facilitates adaptive selection of features and parameters for each sub-task, which contributes to the increase of overall accuracy. Lastly, the framework performs annotation based on the limited set of labeled samples.

In our experiments we demonstrated that the proposed framework outperforms one-step classification by about 10% to 15%. We developed several variants of the proposed framework: 1) we implemented various semi-supervised methods to facilitate decision generation; 2) we compared both manual and automatic feature selection strategies; and 3) we implemented two annotation strategies to annotate concepts. The semi-supervised methods include semi-supervised similarity-based clustering based on K-means and agglomerative clustering, and probabilistic clustering using combination of Gaussian Mixture Model and Expectation maximization. We demonstrated that the probabilistic clustering methodoutperforms similarity-based clustering by up to 5%. Next, we implemented both manual and automatic feature selection at the level of individual experts and demonstrated that while manual feature selection leads to slightly more accurate results, the two variants of the proposed framework generate comparable results. Further, we demonstrated that the full disambiguation of patterns leads to tanhe increase in accuracy by about 7% as compared to the partial disambiguation of patterns.

9. 1. 4 Ontology-based Concepts Disambiguation

We proposed a novel method to perform disambiguation of concepts based on the domain ontology. We pose this problem as Metric Labeling Problem and solve it using Integer Linear Programming. In this method, we exploit ontological relationships and represent them as constraints, while the automatically generated confidence values contribute to the cost function. The goal is to minimize the cost function and, thus, to find the most optimal solution in accordance to both the domain knowledge and the automatically generated judgments. There are several advantages of the proposed method. First, unlike statistical learning techniques, the proposed method does not require a large dataset to perform concept disambiguation. This is especially important for arts collections, where datasets are limited. Second, it is able to handle a large number of concepts. Third, the proposed method relies on the consistent domain knowledge and is robust to the large variety of arts images. Lastly, this method naturally incorporates incomplete annotations, which are often available online, into concept disambiguation process.

In our experiments with medium-size collection of paintings we demonstrated that the proposed method outperforms the widely used majority vote technique by up to 15%. We showed that the proposed method consistently improves precision rates by a minimum of 55% for both ambiguous and non-ambiguous data samples used for concept disambiguation. We also demonstrated the use of this method for concept disambiguation of collections with incomplete online annotations. This method successfully employs incomplete annotations within the disambiguation process. Similar to the setting without online annotations, it generates superior results as compared to the majority vote strategy.

9. 2 Future Work

In our future work, we would like to enhance and extend the existing framework in several directions. First, we would like to utilize the proposed framework for the annotation of abstract-level concepts such as *warm, cold, expressive, rational, gestural* and others. In this thesis, we briefly touched on this topic and demonstrated that the proposed framework performs successful annotation of a small subset of abstract-level concepts. We further aim to extend the set of abstract-level concepts and apply the proposed framework for their annotation. Further, we would like to extend the proposed framework with other visual-level concepts such as composition and aspect information.

Second, the proposed framework utilizes the transductive multi-expert learning approach as discussed in Chapters 7 and 8. In this approach we perform the model selection step, which searchers for the best-performing model by varying the model parameters and the feature subset. We aim to further extend the model selection step and preprocess the pool of models by varying the feature subsets, classification methods used and their parameters for the selection of the best-performing model. This will facilitate better approximation of the data distribution in the semi-supervised model and lead to improved accuracy of annotation.

Lastly, we aim to focus on the application of the proposed framework to the World Wide Web. First, we aim to demonstrate how the proposed ontology-based disambiguation method facilitates full annotation of the partially annotated image collections that are widely available online. Second, we aim to exploit methodologies that relate the three-level ontology of the artistic concepts to the existing arts-oriented ontologies. This will facilitates the publishing of the annotated collection online and their integration with the existing online museum collections and navigational tools. Third, we consider an online social network scenario, where the users are offered to discuss not only visual content of paintings but also their symbolic meaning. We aim to extract and represent the user knowledge as concept ontology and exploit this ontology within the proposed framework for the annotation of artistic concepts.

We also believe that the proposed framework is general and can be extended to other domains such as personal media and news media annotation tasks, where the concept ontology is available. We plan to extend our framework to these tasks, especially with respect to utilizing Web knowledge and social tagging information.

Appendix 1. Software Tools

Implemented	Implemented By	Platform	Available at
CIE L*u*v histogram	Marchenko Yelizaveta	Matlab 7.0.1, Windows XP	N/A
Major colors with account for perceptual similarity	Chua Tat-Seng (C++), adopted by Marchenko Yelizaveta (Matlab)	Matlab 7.0.1, Windows XP	N/A
Color coherence vector	Marchenko Yelizaveta	Matlab 7.0.1, Windows XP	N/A
Support Vector Machine	Chakrabartty, S.	C++, Unix	http://bach.ece.jhu.edu/svm/ginisvm/
Wavelet-based, statistical and model-based texture features	Marchenko Yelizaveta	Matlab 7.0.1, Windows XP	N/A
Gabor texture features	Wei Ying Ma	Matlab 7.0.1, Windows XP	http://vision.ece.ucsb.edu/texture/software/
Multi-expert annotation framework	Marchenko Yelizaveta	Matlab 7.0.1, Windows XP	N/A
Feature and model selection	Marchenko Yelizaveta	Matlab 7.0.1, Windows XP	N/A
Distance-based clustering	Marchenko Yelizaveta	Matlab 7.0.1, Windows XP	N/A
Hierarchical clustering	Marchenko Yelizaveta	Matlab 7.0.1, Windows XP	N/A
GMM and Expectation Maximization	R. Collobert	C++, Windows/Unix	http://www.torch.ch/
Ontology-based Concept Disambiguation	Marchenko Yelizaveta	Matlab 7.0.1, Windows XP	N/A

Table A.1. The list of software tools used in this thesis

Bibliography

Abas, F. S., Martinez, K. Craquelure Analysis for Content-Based Retrieval. In Proceedings of 14th International Conference on Digital Signal Processing 1, pp. 111-114, 2002.

Agrawala, M., Durand, F., Gooch, B., Interrante, V., Ostromoukhov, V., Zorin, D. Perceptual and Artistic Principles for Effective Computer Depiction. Course Notes for SIGGRAPH 2002.

Archer A., Fakcharoenphol J., Harrelson C., Krauthgamer R., Talvar K., and Tardos E., Approximate classification via earthmover metrics, Proceedings of the 15th Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 1079-1087, 2004.

Art & Architecture Thesaurus. Getty Research Institute, 2000.

Arnheim. Art and visual perception: A psychology of the creative eye, University of California Press, 1954.

Asinger C., Kammerer P., Zolda E., Tatzer P. Classification of Color Pigments in Hyperspectral Images. In Proc. of the 10th Computer Vision Winter Workshop (CVWW), pp. 205-214, 2005.

Aslandogan Y. Alp, Their C., Yu C. T., Zou J., Rishe N. "Using Semantic Contents and WordNet in Image Retrieval", SIGIR, pp. 286-295, 1997.

Bagdazian R. Fourier Coding of Image Boundaries. IEEE Trans. PAMI-6(1), pp. 102-105, 1984.

Balcan,M.-F., Blum, A., & Yang, K. Co-training and expansion: Towards bridging theory and practice. In L. K. Saul, Y. Weiss and L. Bottou (Eds.), Advances in neural information processing systems (17), 2005.

Baluja, S. Probabilistic modeling for face orientation discrimination: Learning from labeled and unlabeled data. Neural Information Processing Systems, 1998.

Barnard K., Duygulu P., Forsyth D. "Clustering Art", ComputerVision and Pattern Recognition, pp. 434-439, 2001

Barnard, K., Duygulu, P., de Freitas, N., Forsyth, D., Blei, D., Jordan, M. I.: Matching words and pictures. Journal of Machine Learning Research (3), pp.1107-1135, 2003.

Baum, L. An inequality and associated maximization techniques in statistical estimation of probabilistic functions of Markov processes. Inequalities (3), pp. 1-8, 1972.

Bennett, K. and Demiriz, A. Semi-supervised support vector machines. Advances in Neural Information Processing Systems, 11, pp. 368–374, 1999.

Berezhnoy, I., Postma E. J., Herik, v. D., Digital analysis of van Gogh's complementary colours, In Proc. of 16th Belgian-Dutch Conf. on Artificial Intelligence (BNAIC), p.28-52, 2003.

Berkhin, P. A Survey of clustering data mining techniques. In Grouping Multidimensional Data: Recent Advances in Clustering, pp. 25-71, 2006

Besag, J. Spatial interaction and statistical analysis of lattice systems. Journal of the Royal Statistical Society 36 (3), pp. 192-236, 1974.

Bilenko, M., Basu, S. and Mooney, R.J. Integrating constraints and metric learning in semi-supervised clustering, Proc. of the 21st Int. Conf. on Machine Learning (ICMl), p.11, 2004.

Blei D., Jordan M. I. Modeling annotated data. In Proceedings of the 26th Intl. ACM SIGIR Conf., pp. 127.134, 2003

Blum, A., Langley, P. Selection of relevant features and examples in machine learning. Artificial Intelligence, pp. 245-271, 1997.

Blum, A. and Mitchell, T. Combining labeled and unlabeled data with co-training. COLT: Proceedings of the Workshop on Computational Learning Theory, pp. 92-100, 1998.

Blum, A. and Chawla, S. Learning from labeled and unlabeled data using graph min-cuts. Proc. 18th International Conf. on Machine Learning, pp.19-26, 2001.

Blum, A., & Langford, J. PAC-MDL Bounds, In Proceedings of the Sixteenth Annual Conference on Computational Learning Theory, pp. 344-357, 2003.

Breen C., Khan L., Ponnusamy A., and Wang L. Ontology-based Image Classification Using Neural Networks, Proc. of SPIE Internet Multimedia Management Systems III, pp. 198-208, 2002.

Brandt, S., Laaksonen, J., Oja, E. Statistical Shape Features in Content-Based Image Retrieval. In Proceedings of 15th International Conference on Pattern Recognition (ICPR), pp.187-198, 2000.

Brilliant, R. How an art historian connects art objects and information. Library Trenak (37), pp. 120-129, 1988.

Brown, M. S., Seales, W. B. "The Digital Atheneum: New Approaches for Preserving, Restoring and Analyzing Damaged Manuscripts" Proceedings of the First ACM/IEEE-CS Joint Conference on Digital Libraries, pp. 437-443, 2001.

Canaday J. Mainstreams of Modern Art, Saunders College Publishing, 1981.

Canny J. A Computational Approach to Edge Detection, IEEE PAMI (8), pp.679-698, 1986.

Carneiro G., Vasconcelos N. Formulating Semantic Image Annotation as a Supervised Learning Problem, In Proceedings of CVPR (2), pp. 163-168, 2005.

Carson C., Belongie S., Greenspan H., Malik J. Blobworld. Image segmentation using expectation maximization and its application to image querying, IEEE PAMI 24(8), pp. 1026-1038, 2002.

Castelli, V., & Cover, T. The exponential value of labeled samples, Pattern Recognition Letters (16), pp.105–111, 1995.

Castelli, V., & Cover, T. The relative value of labeled and unlabeled samples in pattern recognition with an unknown mixing parameter. IEEE Transactionson Information Theory (42), pp.2101–2117, 1996.

Chakrabartty, S. and Cauwenberghs, G. Forward Decoding Kernel Machines: A Hybrid HMM/SVM Approach to Sequence Recognition, IEEE Int Conf. On Pattern Recognition, pp.13-27, 2002.

Chang E., Li B. MEGA---the maximizing expected generalization algorithm for learning complex query concepts. ACM Transactions on Information Systems 21 (4), pp.347-382, 2003.

Chang S-F, Eleftheriadis A, McClintock R. Next-generation content representation, creation and searching for new media applications in education. IEEE Proceedings Special Issue on Multimedia Signal Processing (86), pp. 884-904, 1998.

Chang S.-F. The Holy Grail of Content-Based. IEEE MultiMedia 9(2), pp. 6-10, 2002.

Chang, S-K. Hsu, A. Image information systems: Where do we go from here? IEEE Transactions on Knowledge and Data Engineering, 4(5), pp. 431-442, 1992.

Chekuri C., Khanna S., Naor J., and Zosin L., “Approximation algorithms for the metric labeling problem via a new linear programming formulation,” in 12th Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 109–118, 2001

Chen C.-C., Del Bimbo A., Amato G., Boujemaa N., Bouthemy P., Kittler J., Pitas I., Smeulders A., Kirk A., Kiernan K., Li C.-S., Wactlar H., Wang J.-Z. Report of the Delos-Nsf Working Group on Digital Imagery for Significant Cultural and Historical Materials, 2003.

Chen, H. An analysis of image queries in the field of art history. Journal of the American Society for Information Science and Technology, 52(3), pp. 260-273, 2001.

Chen C.-C., Del Bimbo A., Amato G., Boujemaa N., Bouthemy P., Kittler J., Pitas I., Smeulders A., Kirk A., Kiernan K., Li C.-S., Wactlar H., Wang J.-Z. Report of the Delos-Nsf Working Group on Digital Imagery for Significant Cultural and Historical Materials, 2003.

Chua T.-S., Lim S.-K., Pung H.-K.. ”Content-based retrieval of segmented images”. ACM MM, pp. 211 – 218, 1994.

Chua T. S., Low W. C., Chu C. H. Relevance Feedback Techniques for Color-based Image Retrieval. Multimedia Modelling, pp. 24-31, 1998.

Cleverdon, C. W., Mills, J. & Keen, E. M. Factors determining the performance of indexing systems. Cranfield, UK: Aslib Cranfield Research Project, College of Aeronautics. (Volume 1:Design; Volume 2: Results), 1966.

Coggins J.M., Jain A.K. A Spatial Filtering Approach to Texture Analysis. Pattern Recognition Letters 3(3), pp. 195-203, 1985.

Comaniciu D., Meer P. Mean Shift Analysis and Applications. IEEE Int’l Conf. Computer Vision, pp. 1197–1203, 1999.

Corridoni, J. M., Del Bimbo, A., Pala, P. Retrieval of Paintings using Effects Induced by Color Features. In Proceedings of International Workshop on Content-Based Access of Image and Video Databases, p. 2, 1998.

Cortelazzo G. M., Mian G. A., Vezzi G., Zamperoni P. Trademark shapes description by string-matching techniques. Pattern Recognition 27 (8), pp. 1005-1018, 1994.

Cox, I. J, Miller, M., Minka, T., Papathomas, T. V., Yianilos, P. N, The Bayesian Image Retrieval System, PicHunter: Theory, Implementation and Psychophysical Experiments, IEEE Transactions on Image Processing -- special issue on digital libraries 9(1), pp. 20-37, 2000.

Cozman, F., Cohen, I., & Cirelo, M. (2003). Semi-supervised learning of mixture models. ICML-03, 20th International Conference on Machine Learning.

Dara, R., Kremer, S., & Stacey, D. (2002). Clustering unlabeled data with SOMs improves classification of labeled real-world data. Proceedings of the World Congress on Computational Intelligence (WCCI).

Dave R. Characterization and Detection of Noise in Clustering. Pattern Recognition Letters (12), 657-664, 1991.

Davis L. S. Shape matching using relaxation techniques. IEEE Trans. Pattern Anal. Machine Intelligence (1), pp. 60-72, 1979.

DCMI Type Vocabulary. DCMI Recommendation. Dublin core, 2006. Available at http://dublincore.org/documents/dcmi-type-vocabulary/

Demiriz, A., Bennett, K., & Embrechts, M. (1999). Semi-supervised clustering using genetic algorithms. Proceedings of Artificial Neural Networks in Engineering.

Demiriz A., Bennett K. Optimization approaches to semi-supervised learning. Complementarity: Applications,Algorithms and Extensions, volume 50 of Applied Optimization, chapter 1, pp. 1-19.Kluwer, 2000.

Dempster, A., Laird, N., & Rubin, D. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B., 1977.

Derbeko, P., El-Yaniv, R., Meir, R. Explicit learning curves for transduction and application to clustering and compression algorithms. J. Artificial Intelligence Res. (JAIR) 22, pp. 117-142, 2004.

Devijner, P. A, Kittler, J. Pattern Recognition: A Statistical Approach, Prentice-Hall, 1982.

Djeraba, C. Content-Based Multimedia Indexing and Retrieval. In IEEE Multimedia 9(2), pp.18-22, 2002.

Dong A., Li H. Multi-ontology Based Multimedia Annotation for Domain-specific Information Retrieval. IEEE International Workshop on Multimedia Technology and Ubiquitous Computing, 2006.

Dorai, C., Venkatesh, S. Bridging the Semantic Gap with Computational Media Aesthetics. IEEE MultiMedia 10(2), pp.15-17, 2003.

Duda R. O., Hart P. E., Stork D. G., Pattern Classification (2nd ed), Wiley, 2000.

Duygulu, K. Barnard, N. de Freitas, and D. Forsyth. “Object Recognition as Machine Translation: Learning a Lexicon for a Fixed Image Vocabulary,” Proc. ECCV, 2002.

Eakins, J. P. “Content-based Image Retrieval. A report to the ISC Technology Applications Programme”. Institute for Image Data Research, University of Northumbria at Newcastle, Technical Report, January 1999.

El-Yaniv, R., and Gerzon, L. Effective Transductive Learning via PAC-Bayesian Model Selection. Technical Report CS-2004-05, IIT, 2004.

Elworthy, D. Does Baum-Welch re-estimation help taggers? Proceedings of the 4th Conference on Applied Natural Language Processing, pp. 53–58, 1994.

Enser, P.G.B. Query analysis in a visual information retrieval context. JDocument & Text Management 1, pp. 45-52, 1993.

Eronen A. Musical instrument recognition using ICA-based transform of features and discriminatively trained HMMs. Proceedings of the Seventh International Symposium on Signal Processing and its Applications (ISSPA), pp. 133-136, 2003.

Faloutsos C., FlicknerM., Niblack W., Petkovic D., Equitz W., Barber R Efficient and Effective Queryingby Image Content, Technical Report, IBM Research Report, 1993.

Fan J., Gao Y., Luo H., Xu G.: Statistical modeling and conceptualization of natural images. Pattern Recognition 38(6), pp. 865-885, 2005.

Fan J., Luo H., Gao Y. Learning the semantics of images by uing unlabeled samples, IEEE International Computer Vision and Pattern Recognition, 2005.

Fellbaum G. WordNet An Electronic Lexical Database, MIT PRESS, 2001.

Feng H., Shi R. and Chua T.-S. A Bootstrapping Framework for Annotating and Retrieving WWW Images. ACM Multimedia, pp. 960-967, 2004a.

Feng H., Chua T.-S. A Learning-based Approach for Annotating Large On-Line Image Collection. In Proceedings of International Conference on Multimedia Modeling, 2004b.

Fidel, R. The image retrieval task: Implications for the design and evaluation of image databases. The New Review of Hypermedia and Multimedia, 3, 181-199, 1997.

Flickner, M.D., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, G., Lee, D., Petkovic, D., Steele, D., and Yanker, P. Query by image and video content: The QBICsystem. Computer, 28(9):23–32, September 1995.

Friedman J. H., An overview of predictive learning and function approximation, From Statistics to Neural Networks, Springer Verlag, NATO/ASI, 1-61,1994.

Forsyth D. and Fleck M.. Body Plans. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 678–683, 1997.

Fung, G., &Mangasarian, O. Semi-supervised support vector machines for unlabeled data classification (Technical Report 99-05). Data Mining Institute, University of Wisconsin Madison, 1999.

Furht B. Handbook of Internet Computing. CRC Press 2000.

Garber, S. R., Grunes, M. B. The art of search: A study of art directors. CHI Human Factors in Computing Systems, pp.157-163, 1992.

Getty Research Institute. Getty Vocabulary Program. Art & Architecture Thesaurus, Februrary 2000. Available at http://shiva.pub.getty.edu/aat_browser/

Gluskman H. A. Multicategory classification of patterns represented by high-order vectors of multilevel measurements. IEEE Tran son Computers (20), 1593–1598, 1971.

Goldman, S. and Zhou, Y. Enhancing supervised learning with unlabeled data. Proc. 17th International Conf. on Machine Learning, pp. 327–334, Morgan Kaufmann, 2000.

Gonzalez R. C., Woods R. E. "Digital Image Processing", Addision Wesley, 1992.

Grady, L., Funka-Lea, G. Multi-label image segmentation for medical applications based on graph-theoretic electrical potentials. ECCV 2004 workshop, 2004.

Greenberg, J. Intellectual Control of Visual Images: A Comparison Between the Art and Architecture Thesaurus and the Library of Congress Thesaurus for Graphic Materials. Cataloging & Classification Quarterly 16 (1), pp. 85-101, 1993.

Grira, N., Crucianu, M. and Boujemaa, N. Unsupervised and semisupervised clustering: a brief survey. in 'A Review of Machine Learning Techniques for Processing Multimedia Content', Report of the MUSCLE European Network of Excellence (FP6), 2004.

Grira, N., Crucianu, M. and Boujemaa, N. Semi-supervised image database categorization using pairwise constraints. International Conference on Image Processing (ICIP'05), Genoa, Italy, september 2005.

Gruber T. R. "A Translation Approach to Portable Ontology Specifications." Knowledge Acquisition, 5(2), pp. 199-220, 1993.

Gyftodimos, E., Flach, P. Hierarchical Bayesian networks: an approach to classification and learning for structured data. Proceedings of the Work-in-Progress Track at the 13th International Conference on Inductive Logic Programming, pp. 12–21, 2003.

Haering H., Myles Z., Lobo N.. Locating Dedicuous Trees. In Workshop in Content-based Access to Image and Video Libraries, pp. 18–25, 1997.

Hall R. Illumination and color in computer generated imagery, Springer-Verlag New York, Inc., New York, NY, 1988.

Hanbury A. , Kammerer P. , and Zolda E. Painting Crack Elimination Using Viscous Morphological Reconstruction. In Proceedings of the 12th Intl. Conf. on Image Analysis and Processing, ICIAP2003, pp. 226-231, 2003.

Hartigan, J.A. Clustering Algorithms, NY: Wiley, 1975.

Hartigan, J.A., and Wong, M.A. Algorithm AS136: A k-means clustering algorithm. Applied Statistics (28), pp. 100-108, 1979.

Hastings, S. K. Query categories in a study of intellectual access to digitized art images. Proceedings of the 58th Annual Meeting of the American Society for Information Science, 1995.

Herik, H.J. van den, Postma, E.O. Discovering the Visual Signature of Painters. In Future Directions for Intelligent Systems and Information Sciences, 129-147, 2000.

Hertzmann, A. "Painterly Rendering with Curved Brush Strokes of Multiple Sizes. SIGGRAPH 98 Conference Proceedings". pp. 453-460. Orlando, Florida. July, 1998.

Ho T. K., Hull J. and Srihari S. N. Decision combination in multiple classifer systems. IEEE Transactions on Pattern Analysis and Machine Intelligence, 16(1):66-75, January 1994

Hollink L., Schreiber A., Wielemaker J., Wielinga B. "Semantic Annotation of Image Collections," In Proceedings of the Workshop on Knowledge Capture and Semantic Annotation (KCAP), 2003.

Holt, B., Hartwick, L., and Vetter, S. “Query by Image Content, the QBIC Project's Applications at Davis's Art and Art History Departments”. Visual Resources Association Journal, 22, no. 2, Summer, 1995.

Hsu W., Chua T.-S., Pung H K. An Integrated Color-Spatial Approach to Content-Based Image Retrieval. ACM Multimedia, pp. 305-313, 1995.

Hu B., Dasmahapatra S., Lewis P., Shadbolt N., “Ontology-Based Medical Image Annotation with Description Logics”, IEEE ICTAI’03, November 03-05, 2003.

Huang, P.W., Jean, Y.R., Using 2d C+ Strings As Spatial Knowledge Representations For Image Database Systems, Pattern Recognition (27), pp. 1249-1257, 1994.

Hyvönen E., Saarela S., Viljanen K. “Ontology Based Image Retrieval”. In Proceedings of WWW, 2003.

Icoglu O., Gunsel B., Sariel S. Classification and Indexing of Paintings Based on Art Movements. EUSIPKO, 2004.

Imam, I.F., Michalski, R.S., and Kerschberg, L. “Discovering Attribute Dependence in Databases by Integrating Symbolic Learning and Statistical Analysis Techniques”, AAAI Workshop on Knowledge Discovery in DB, 1993.

Itten J. “The Art of Color”, Reinhold Pub. Corp., NY, 1961.

Jain A. K., Vailaya A. Image Retrieval using Color and Shape. Second Asian Conference on Computer Vision, Singapore, pp. 529-533, 1995.

Jain A. K., Vailaya, A. Shape-Based Retrieval: A Case Study With Trademark Image Databases. Pattern Recognition 31(9), pp.1369--13990, 1998.

Jain R., Kasturi R., Schunck B.G. Machine Vision, McGraw-Hill, 1995.

Jain R. NSF Workshop on Visual Information Management Systems: Workshop Report. Storage and Retrieval for Image and Video Databases (SPIE), pp. 198-218, 1993.

Jeon J., Lavrenko V. and Manmatha R. Automatic Image Annotation and Retrieval using Cross-Media Relevance Models. In Proceedings of the 26th Intl. ACM SIGIR Conf., pp. 119-126, 2003.

Jiang S., Huang T.,Gao W. “An Ontology-based Approach to Retrieve Digitized Art Images”. Web Intelligence, 131-137, 2004.

Joachims, T. Transductive inference for text classification using support vector machines. Proc. 16th International Conf. on Machine Learning, pp. 200–209, 1999.

Joachims, T. Transductive learning via spectral graph partitioning. Proceedings of ICML-03, 20th International Conference on Machine Learning, 2003.

Jorgensen, C. Attributes of images in describing tasks. Information Processing and Management, 34(2), 161-174, 1998.

Jorgensen C., Jaimes, A. , Benitez, A. B., Chang, S.-F., "A conceptual Framework and Research for Classifying Visual Descriptors", invited paper, Journal of the American Society for Information Science (JASIS), special issue on "Image Access: Bridging Multiple Needs and Multiple Perspectives", 2001.

Kammerer P. Pose Estimation and Comparison of Painted Portraits using a 3D Head Model. In Tomáš Svoboda, editor, Proceedings of the Czech Pattern Recognition Workshop 2000, pages 173-178, 2000.

Kammerer P., Sablatnig R and Zolda E. Head Pose Estimation in Painted Portraits used for Comparison. In S. Scherer, editor, Computer Vision, Computer Graphics and Photogrammetry - a Common Viewpoint, Proc. of the 25th Workshop of the Austrian Association for Pattern Recognition (OEAGM), volume 147, pp. 127-134, 2001.

Kammerer P. , Lettner M. , Zolda E. and Sablatnig R. Identification of Drawing Tools by Classifiation of Textural and Boundary Features of Strokes. Pattern Recognition Letters, Special Issue, 2004.

Kaplan L. M. and Kuo C.-C. J., “Texture roughness analysis and synthesis via extended self-similar (ESS) model,” IEEE Trans. Pattern Anal. Machine Intell (17), 1043–1056, 1995.

Keren, D. “Recognizing image style and activities in video using local features and naive Bayes”, Pattern Recognition Letters, vol. 24(16), pp. 2913-2922, 2004.

Kittler J, Hatef M. Improving recognition rates by classifier combination. 5th Int Workshop on Frontiers of Handwriting Recognition, 81–102, 1996.

Klapuri, A., Eronen, A., Astola, J. Analysis of the meter of acoustic musical signals. IEEE Trans. Audio, Speech, and Language Processing, 14(1), 2006.

Klein, D., Kamvar, S. D., & Manning, C. From instance-level constraints to space-level constraints: Making the most of prior knowledge in data clustering. Proceedings of ICML, 307-314, 2002.

Kleinberg J. and Tardos E. “Approximation algorithms for classification problems with pair-wise relationships: metric labeling and Markov random fields,” Journal of the ACM, vol. 49, pp. 616–630, 2002.

Kohavi R., John G. Wrappers for feature subset selection. Artificial Intelligence Journal, 97(1–2), pp. 273–324, 1997.

Kropatsch W. G., Eder M. and Kammerer P. Finding strokes of the brush in portrait miniatures. In Franc Solina and Walter G. Kropatsch, editors, Visual Modules, Proc. of 19th ÖAGM and 1st SDVR Workshop, pp. 257-265, 1995.

Kuncheva, L. I. Combining Pattern Classifiers:Methods and Algorithns. John Wiley & Sons, Inc, Hoboken, NJ, 2004.

Lavrenko V., Manmatha R., Jeon J., “A Model for Learning the Semantics of Pictures”, Neural Information Processing Systems, 2003.

Lay, J., Guan, L., "Retrieval of color artistry by color concepts," IEEE Trans.on Image Processing, vol. 13, no. 2, pp. 326-339, March 2004.

Layne S. S. Some issues in the indexing of images. Journal of the American Society for Information Science, 45(8), 583-588, 1994.

Lavrenko V., Manmatha R., Jeon J. A Model for Learning the Semantics of Pictures. NIPS, 2003.

Lazzari M. R., Lee C. “Art and Design Fundamentals”, Van Nostrand Reinhold, 1990.

Lehmann, E. Elements of Large Sample Theory. New York City, Springer, 1999.

Lettner M. , Kammerer P. and Sablatnig R. Texture Analysis of Painted Strokes. In W. Burger and J. Scharinger, editors, 28th Workshop of the Austrian Association for Pattern Recognition (OAGM/AAPR), pp. 269-276, 2004.

Lettner M., Sablatnig R. Texture Analysis for Stroke Classification in Infrared Reflectogramms. In 14th Scandinavian Conference on Image Analysis, pp. 459-469, 2005.

Lew M., Sebe N., Djeraba C., Jain R. Content-based Multimedia Information Retrieval: State-of-the-art and Challenges, ACM Transactions on Multimedia Computing, Communication, and Applications, Vol 2 (1), pp. 1-19, 2006.

Lewis, P. H., Martinez, K., Abas, F. S., Ahmad Fauzi, M. F., Addis, M., Lahanier, C., Stevenson, J., Chan, S. C. Y., Mike J., B. and Paul, G. An Integrated Content and Metadata based Retrieval System for Art. IEEE Transactions on Image Processing 13(3), pp. 302-313, 2004.

Li J., Wang J. Z. Studying Digital Imagery of Ancient Paintings by Mixtures of Stochastic Models, IEEE Transactions on Image Processing, vol. 13 (3), 2004.

Li Y., Shapiro L. Consistent line clusters for building recognition in CBIR. In International Conference on Pattern Recognition (3), pp. 952–956, 2002.

Li W., Sun M. Semi-supervised Learning for Image Annotation Based on Conditional Random Fields. CIVR, 463-472, 2006.

Library of Congress. Thesaurus for Graphic Materials I: Subject Terms. (February 2000). Available at http://lcweb.loc.gov/rr/print/tgm1.

Ma W.-Y., Manjunath B. S. Texture-Based Pattern Retrieval from Image Databases. Multimedia Tools Appl. 2(1), pp. 35-51, 1996.

Ma W.-Y., Manjunath B. S. NeTra: A Toolbox for Navigating Large Image Databases. Int. Conf. On Image Processing (1), pp. 568-571, 1997a.

Ma W.-Y., Manjunath B. S. Edge Flow: A Framework of Boundary Detection and Image Segmentation. CVPR, pp. 744-749, 1997.

Maddage, N.C., Xu, C. S., Lee, C. H., Kankanhalli, M., Tian, Q. Statistical Analysis of Musical Instruments. Proc. Third IEEE Pacific-Rim Conference on Multimedia PCM2002, 2002.

Mallat S. A theory for multi-resolution signal decomposition: the wavelet representation, IEEE PAMI (11), 674-693, 1989.

Mandelbrot B. B, The Fractal Geometry of Nature. San Francisco, CA: Freeman, 1982.

Manjunath B. S., Ma W. Y., “Texture features for browsing and retrieval of image data,” IEEE Trans. Pattern Anal. Machine Intell (18), 837–842, 1996.

Manjunath B.S., Salembier P., Sikora T. Introduction to MPEG-7: Multimedia Content Description Interface. Wiley & Sons, 2002.

Marciniak T., Strube M. Beyond the pipeline: discrete optimization in NLP.
In: Proceedings of the 9th Conference on Computational Natural Language Learning, Ann Arbor, Michigan, June 29-30, 2005, pages 136-143.

Marchenko Y., Chua T.-S., Aristarkhova I., Jain R. Representation and Retrieval of Paintings based on Art History Concepts. IEEE Int'l Conf. on Multimedia and Expo (ICME), 2004.

Marchenko Y., Chua T.-S., Aristarkhova I., Analysis of paintings using Color Concepts. IEEE Int'l Conf. on Multimedia and Expo (ICME), 2005.

Marchenko Y., Chua T.-S., Jain R., Semi-supervised Annotation of Brushwork in Painting Domain using Serial Combinations of Multiple Experts, *ACM Multimedia,* 2006.

Marchenko Y., Chua T.-S., Jain R., Transductive Inference Using Multiple Experts for Brushwork Annotation in Paintings Domain, *ACM Multimedia,* 2006.

Marchenko Y., Chua T.-S., Jain R., Ontology-Based Annotation of Paintings using Transductive Inference Framework, *The International MultiMedia Modeling Conference, Best Paper Award,* 2007.

Maeireizo, B., Litman, D., & Hwa, R. (2004). Co-training for predicting emotions with spoken dialogue data. The Companion Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL).

McBride B., BrickleyD., Guha R.V., RDF Vocabulary Description Language 1.0: RDF Schema, W3C Recommendation, 2004.

McLachlan, G.J. and Basford, K.E. Mixture Models, New York: Marcel Dekker, Inc., 1988.

Mehrotra S., Rui Y., Ortega M., Thomas S. Huang: Supporting Content-based Queries over Images in MARS. ICMCS, pp. 632-633, 1997.

Melzer, T., Kammerer, P., Zolda E. Stroke detection of Brush Strokes in Portrait Miniatures using Semi-Parametric and a Model-Based Approach. In Proceedings of International Conference on Pattern Recognition, 1998.

Mezaris V., Kompastsiaris I., Strintzis M. G., “An Ontology Approach to Object-Based Image Retrieval”, ICIP 2003.

Miller D. J., Uyar H. S. A mixture of expert classifiers with learning based on both labelled and unlabelled data. In M. Mozer, M. I. Jordan, and T. Petsche, editors, Advances in Neural Information Processing Systems 9, pp. 571-577. MIT Press, Cambridge, MA, 1997.

Miller D. J., Browning J. A Mixture Model and EM-Based Algorithm for Class Discovery, Robust Classification, and Outlier Rejection in Mixed Labeled/Unlabeled Data Sets. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25(11), pp. 1468-1483, 2003.

Minka, T., Picard, R. Interactive learning using a "society of models". In Proceedings IEEE Conference.on Computer Vision and Pattern. Recognition, 1996.

Minsky M. A framework for representing knowledge. The Psychology of Computer Vision, 1975.

Mitchell T. Machine Learning, McGraw Hill, 1997.

Mitchell, T. (1999). The role of unlabeled data in supervised learning. Proceedings of the Sixth International Colloquium on Cognitive Science, 1999.

Monay F., Gatica-Perez D. On Image Auto-Annotation with Latent Space Models. Proceedings of the 7th ACM Multimedia, 2003.

Mori Y., H. Takahashi H., Oka R. Image-to-word transformation based on dividing and vector quantizing images with words. In MISRM'99 First International Workshop on Multimedia Intelligent Storage and Retrieval Management, 1999.

Mylonas Ph., Athanasiadis Th. and Avrithis Y.. Image Analysis Using Domain Knowledge and Visual Context. In proceedings of 13th International Conference on Systems, Signals and Image Processing (IWSSIP), 2006.

Multiple Classifier Systems: First/Second/Third/Fourth/Fifth International Workshop (MCS-2000/2001/2002/ 2002/2003/2004) Springer-Verlag GmbH, 2000-2004.

Munsell A. H. The Atlas of the Munsell Color System, Boston, 1915.

Murtagh, F. (1983), "A Survey of Recent Advances in Hierarchical Clustering Algo-. rithms," Computer Journal (26), 354-359.

Murphy K., Torralba A., Freeman W. T. Using the forest to see the trees: a graphical model relating features, objects, and scenes. *Adv. Neural Inf. Process. Sys.*, 16, 2003.

Nigam, K. and Ghani, R. Analyzing the effectiveness and applicability of co-training. Ninth International Conference on Information and Knowledge Management, pp. 86–93, 2000.

Nigam, K., McCallum, A. K., Thrun, S. and Mitchell, T. Text classification from labeled and unlabeled documents using EM. Machine Learning, 39, pp. 103–134, 2000.

Olson, C.F.: Parallel algorithms for hierarchical clustering. Parallel Computing 21, pp. 1313–1325, 1995.

Ortega M., Rui Y., Chakrabarti K., Porkaew K., Mehrotra S., Huang T. S. Supporting Ranked Boolean Similarity Queries in MARS,” IEEE Trans. Knowledge and Data Eng., vol. 10 (6), pp. 905-925, 1998.

Pan j., Yang H. J., Duygulu P., Faloutsos C. Automatic Image Captioning. In Proceedings of the 2004 IEEE International Conference on Multimedia and Expo (ICME), 2004.

Panofsky, E. Studies in iconology. New York: Harper and Row, 1962.

Parsons L., Haque E., and Liu H., Subspace clustering for high dimensional data: a review, SIGKDD Explor. Newsl. 6 (1), 90-105, 2004.

Pass G., Zabih R., and Miller J. “Comparing Images Using Color Coherence Vectors”, ACM MM, pp. 65-73, 1996.

Pavlidis T. A review of algorithms for shape analysis. Computer Graphics and Image Procesing (7), pp.243–258, 1978.

Petland A. P. Fractal-Based Description of Natural Scenes, IEEE Transactions (9), pp. 661-674, 1984.

Petland A., Picard R. W., Sclaroff S. Photobook: Content-based Manipulation of image database. International Journal of Computer Vision, Vol.18 (3), pp.233-254, 1996.

Petridis K., Bloehdorn S., Saathoff C., Simou N., Dasiopoulou S., Tzouvaras V., Handschuh S., Avrithis Y., Kompatsiaris I. and Staab S. Knowledge Representation and Semantic Annotation of Multimedia Content. IEEE Proceedings on Vision Image and Signal Processing, Special issue on Knowledge-Based Digital Media Processing, Vol. 153(3), pp. 255-262, 2006.

Pudil P, Novovicova J, Blaha S. Multistage pattern recognition with reject option,1th IAPR ICPR, 92–95, 1992.

Pumphrey, R. Elements of Art. Upper Saddle River, NJ : Prentice Hall, 1996.

Punyakanok V., Roth D., Yih W., and Zimak D.. Semantic role labeling via integer linear. programming inference. In Proceedings of International Conference on Computational Linguistics, 2004.

Rahman A., Fairhurst M. An Evaluation Of Multi-Expert Configurations For The Recognition Of Handwritten Numerals, Pattern Recognition 31(9), 1255-1273, 1998.

Rahman A. F. R, Fairhurst M. C: Serial Combination of Multiple Experts: A Unified Evaluation. Pattern Anal. Appl. 2(4), 292-311, 1999.

Ratsaby, J., & Venkatesh, S. (1995). Learning from a mixture of labeled and unlabeled examples with parametric side information. Proceedings of the Eighth Annual Conference on Computational Learning Theory, 412–417.

Riloff, E., Wiebe, J., &Wilson, T. (2003). Learning subjective nouns using extraction pattern bootstrapping. Proceedings of the Seventh Conference on Natural Language Learning (CoNLL-2003).

Rosenberg, C., Hebert, M., & Schneiderman, H. Semi-supervised self-training of object detection models. Seventh IEEE Workshop on Applications of Computer Vision, 2005.

Roth, D. & W. Yih (2004). A linear programming formulation for global inference in natural language tasks. In Proceedings of the 8th Conference on Computational Natural LanguageLearning, Boston, Mass., May 2-7, 2004, pp. 1–8.

Rui Y., Huang T. S., Ortega M., Mehrotra S. Relevance Feedback: A Power Tool in Interactive Content-Based Image Retrieval. IEEE Tran on Circuits and Systems for Video Technology, Special Issue on Segmentation, Description, and Retrieval of Video Content, vol. 8(5), pp. 644-655, Sept, 1998

Rui S., Feng H., Chua T.-S., Lee C.-H. An Adaptive Image Content Representation and Segmentation Approach to Automatic Image Annotation. Int'l Conference on Image and Video Retrieval (CIVR' 04), pp. 545-554, 2004.

Saathoff C., Petridis K., Anastasopoulos D., Timmermann N., Kompatsiaris I. and Staab S. M-OntoMat-Annotizer: Linking Ontologies with Multimedia Low-Level Features for Automatic Image Annotation. In proceedings of the 3rd European Semantic Web Conference (ESWC), 2006.

Sablatnig R., Kammerer P. and Zolda E. Structural Analysis of Paintings based on Brush Strokes. In Anti-Counterfeiting in Art, IS&T/SPIE's 10th Annual Symposium on Electronic Imaging, 1998.

Saltzman, Benthall A. Art slide retrieval: one library´s solution. "MC Journal: The Journal of Academic Media Librarianship". vol. 8(2), 2002.

Seeger, M. (2001). Learning with labeled and unlabeled data (Technical Report). University of Edinburgh.

Schettini R. Multicolored object recognition and location. Pattern Recognition Letters, vol. 15, pp. 1089--1097, 1994.

Schreiber A.T., Dubbeldam B. “Ontology-based photo annotation”, IEEE Intelligent Systems, 2001.

Schreiber A., Blok I. “A Mini-experiment in Semantic Annotation”. The Semantic Web-ISWC, LNCS 2342, pp. 404-408, 2002.

Schueermann J, Doster W. A decision theoretic approach to hierarchical classifier design. Pattern Recognition; 17(3), 359–369, 1983.

Skounakis M., Craven M. Evidence combination in biomedical natural-language processing. In *BIOKDD*, 2003.

Shi R., Jin W., Chua T.-S. A Novel Approach to Auto Image Annotation Based on Pairwise Constrained Clustering and Semi-Naïve Bayesian Model. International Conference on Multimedia Modeling, pp. 322-327, 2005.

Shi R., Chua T.-S., Lee C.-H., Gao S. Bayesian Learning of Hierarchical Multinomial Mixture Models of Concepts for Automatic Image Annotation. CIVR, pp. 102-112, 2006.

Silva, L., Mastella, L. S., Abel, M., Galante, R. M., De Ros, L. F. Ontology-based approach for visual knowledge: Image annotation and interpretation. Workshop on Ontologies and their Applications, 2004.

Smeulders A., Hardman L. Schreiber G. Integrated Access to Cultural Heritage E-Docs. Proceedings 4th Intl. Workshop on Multimedia Information Retrieval, Juan-les-Pins, 2002.

Smith, J. R., Chang, S.-F. VisualSEEk: a fully automated content-based image query system. ACM Multimedia, 1996.

Smith J. R., Chang S.-F. An Image and Video Search Engine for the World-Wide Web. In Symposium on Electronic Imaging: Science and Technology - Storage & Retrieval for Image and Video Databases V, 1997.

Soo V.-W., Lee C.-Y., Yeh J. J., Chen C.-C. “Using sharable ontology to retrieve historical images”, JCDL, pp.197-198, 2002.

Soo V.-W., Lee C.-Y., C.-C. Li, Chen S. L., Chen C.-C. “Automated Semantic Annotation and Retrieval Based on Sharable Ontology and Case-Based Learning Techniques”. JCDL, 2003.

Srikanth M, Varner J., Bowden M. and Moldovan D. Exploiting Ontologies for Automatic Image Annotation”, In Proceedings of the 28th ACM SIGIR, 2005.

Stanchev P., Green D, Jr, Dimitrov B. “High level color similarity retrieval”, Information Theories & Applications, vol. 10(3), pp. 283-287, 2003

Szummer M., Picard R. Indoor-Outdoor Image Classification. In Workshop in Content-based Access to Image and Video Databases, 1998.

Swain M.J.,.Ballard, D.H Color indexing. Intl. Journal of Computer Vision,vol.(1), pp. 11-32, 1991.

Tajudin S., Landgrebe D. Robust Parameter Estimation for Mixture Model. IEEE Trans. Geoscience and Remote Sensing (38), pp. 439-445, 2000.

Tamura, H., Mori, S., Yamawaki, T.: Textural features corresponding to visual perception. IEEE Transactions on Systems, Man, and Cybernetics 8(6), pp. 460–472, 1978.

Tanaka S., Kurumizawa J., Inokuchi S, Iwadate Y. “Composition Analyzer: support tool for composition analysis on painting masterpieces”, Knowledge-Based Systems 13(7-8), pp. 459 - 470, 2000.

Teague, M.R. Image Analysis via the General Theory of Moments, Journal of the Optical Society of America, 70 (8), 920-930, 1979.

Truong B. T., Venkatesh S., Dorai C. “Application of Computational Media Aesthetics Methodology to Extracting Color Semantics in Film”, ACM Multimedia, 2002.

Tsai T.-H., Wu C.-W., Lin Y.-C., Hsu W.-L. Exploiting Full Parsing Information to Label Semantic Roles Using an Ensemble of ME and SVM via Integer Linear Programming. The CoNLL-2005 Shared Task on Semantic Role Labeling

Tuceryan M. and Jain A. K. Texture Analysis. In The Handbook of Pattern Recognition and Computer Vision (2nd Edition), by C. H. Chen, L. F. Pau, P. S. P. Wang (eds.), pp. 207-248, World Scientific Publishing Co., 1998.

Tzanetakis G. and Cook P. Musical Genre Classification of Audio Signals. IEEE Transactions on Speech and Audio Processing, 10(5), July 2002.

Vailaya A., Jain A, Zhang H. On Image Classification: City vs. Landscape. Pattern Recognition (31), pp.1921–1936, 1998.

Vapnik, V. N. (1982). Estimation of Dependences Based on Empirical Data. New York: Springer Verlag.

Vapnik, V. N. (1995). The Nature of Statistical Learning Theory. New York: Springer Verlag.

Vapnik, V. N. (1998). Statistical Learning Theory. New York: Wiley Interscience.

Voorhees, E.M., Harman, D.K. TREC: Experiment and evaluation in information retrieval, Cambridge MA: MIT Press, 2005. ACM SIGIR Forum 20 Vol.40 (1), 2006.

von Ahn, L., Dabbish, L. Labeling images with a computer game Laura Proceedings of the 2004 conference on Human factors in computing systems, pp.319-326, 2004.

Waal, H.v.d. “ICONCLASS: An Iconographic Classification System”, Koninklijke Nederlandse Akademie van Wetenschappen, 1985.

Wagstaff, K., Cardie, C., Rogers, S. and Schroedl, S. Constrained K-means clustering with background knowledge. Proc. of Int’l Conference on Machine Learning (ICML), 2001.

Walton, K. Style and the Products and Processes of Art. In Lang, B.(Ed.), The Concept of Style, pp. 45-66, University of Pennsylvania Press, 1979.

Wang, J., Li, J., Wiederhold, G., SIMPLIcity: Semantics-Sensitive Integrated Matching of Picture Libraries, IEEE PAMI, vol. 23(9), 2001.

Wang J. Z., Li J. Learning-based linguistic indexing of pictures with 2-D MHMMs. Proc. ACM Multimedia, pp. 436-445, 2002.

Wang J. Z, Grieb K., Zhang Y., Chen C.-C., Chen Y., Li J. Machine annotation and retrieval for digital imagery of historical materials, International Journal on Digital Libraries, 6(1), pp. 18–29, 2006.

Wu D., Bennett K., Cristianini N., Shawe-Taylor J. Large margin trees for induction and transduction. In International Conference on Machine Learning, 1999.

Wielinga B., Schreiber G., Wielemaker J., Sandberg J. A. C. “From thesaurus to ontology”. International Conference on Knowledge Capture, 2001.

Xing E. P., Ng A. Y., Jordan M. I., and Russell S.. Distance metric learning, with application to clustering with side-information. In Advances in Neural Information Processing Systems (15), pp. 505–512, 2003.

Yang, Y., Liu, X., A re-examination of text categorization methods, In Proc. of Int'l ACM Conf. on Research and Development in Information Retrieval (SIGIR), pp. 42-49, 1999.

Yang J., Wenyin L., Zhang H., Zhuang Y. “Thesaurus-Aided Approach for Image Retrieval and Browsing”. In Proceedings of 2nd IEEE International Conference on Multimedia and Expo (ICME), pp. 313-316, 2001.

Yarowsky, D. Unsupervised word sense disambiguation rivaling supervised methods. Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics, pp. 189–196, 1995.

Zhang D. S.and Lu G. Content-Based Shape Retrieval Using Different Shape Descriptors: A Comparative Study, IEEE ICME, 2001.

Zhao Q., Miller, D.J.Semisupervised learning of mixture models with class constraints. IEEE International Conference on Acoustics, Speech, and Signal Processing (5), pp. 185-188.

Zhou, Z.-H. and Li, M. Semi-supervised regression with co-training. International Joint Conference on Artificial Intelligence (IJCAI), 2005a.

Zhou, Z.-H. and Li, M. Tri-training: exploiting unlabeled data using three classifiers. IEEE Transactions on Knowledge and Data Engineering, 17, 1529–1541, 2005b.