The Pennsylvania State University

The Graduate School

Eberly School of Science

# POTENTIAL ENERGY DISTANCE BASED IMAGE RETRIEVAL

A Thesis in

Statistics

By

Qi Fang



Submitted in Partial Fullfillment

of the Requirements

for the degree of

Master of Science

August 2013

The thesis of Qi Fang was reviewed and approved* by the following:

Jia Li

Professor of Statistics

Professor of Computer Science and Engineering (by courtesy)

Thesis Advisor

Le Bao

Assistant Professor of Statistics

David Hunter

Professor and Department Head of Statistics

*Signatures are on file in the Graduate School

# Abstract

Due to the large-scale use of digital cameras and easy access to the Internet, the number of images available on the Internet has exploded over the last twenty years and continues to grow. As a result, effective content-based image retrieval approaches are needed to help analyze and understand the large scale of images.

Image retrieval approaches depend greatly on similarity measures. There are several popular similarity measures used for image retrieval, for example, Mallows distance and integrated region matching (IRM). Mallows distance is a metric, but IRM is much faster to compute. This thesis introduces a new measure - potential energy distance for image similarity computation. Potential energy distance is also a metric, but is much faster to compute.

To evaluate the performance of potential energy distance, we conduct experiments to compare potential energy distance with Mallows distance and IRM. In our experiment, we use the MIR Flickr dataset that contains 25,000 images, and we evaluate different similarity measures from accuracy, speed, and robustness perspectives. Experiment results show that potential energy distance performs similarly to Mallows and IRM in accuracy, similarly to IRM but much faster than Mallows in speed, and fairly robust in image alternations.

# Table of Contents

# List of Figures

# List of Tables

# Chapter 1 Introduction

## 1.1. Image Retrieval

Over the past twenty years, the volume of images on the Internet has grown explosively. Due to the invention and general adoption of digital devices, such as cameras and smart phones, the number of pictures generated by common users has increased dramatically. Additionally, the improved speed and ease of Internet access and the popularity of social websites have encouraged common users to share pictures online. As a result, the number of images available online is tremendous and continues to grow. For example, by August 2011, Flickr was hosting more than 6 billion images, and this number has grown steadily. Similarly, approximately 2.5 billion photos are uploaded to Facebook each month, around 50 billion cumulatively.

This rapidly growing volume of images requires significant effort to understand, categorize, and retrieve images, leading to growth and prosperity in the area of image retrieval. Image retrieval, a technology that helps users understand and organize images by content, requires new technology and systems, and involves many researchers from a variety of disciplines. Image retrieval has also encouraged growth in related research areas, such as computer vision, information retrieval, human-computer interaction, machine learning, Web mining, data mining, database systems, statistics, psychology, and so on [1].

As the amount of images increases, the variation of visual and semantic contents of images also increases significantly. The variation of image content and the increase of image size make image computation and understanding more difficult. While these factors provide great

opportunities for image retrieval and understanding, they also challenge the field. Hence, despite efforts to study image retrieval, it is still difficult to interpret images. As a result, efficient image retrieval approaches remain urgent.

Image retrieval approaches can be generally divided into two categories: text-based image retrieval and content-based image retrieval. Text-based approaches rely on textual descriptions of images and retrieve images based on various similarity measures of text descriptions. Content-based approaches analyze images at the pixel level and perform image retrieval based on various similarity measures of visual content. Text-based approaches are generally adopted. Major search engines, such as Google, MSN, and Yahoo, use text-based approaches by extracting filenames, tags, and other text information associated with images. But text-based approaches have shortcomings when retrieving images in some situations, such as when images do not have text descriptions, have text descriptions that are at different abstract levels, or have subjective text descriptions.

To avoid these problems, content-based image retrieval does not rely on text descriptions. Unfortunately, content-based approaches have their own shortcomings because they rely on visual similarity to generate semantic similarity, which results that lack coincidence between visual similarity and semantic similarity due to the semantic gap [2]. Semantic gap exists because there is difference between the information extracted from the visual data and the interpretation that the visual data have for users in specific situations. Images may exist by themselves, thus image signatures only rely on data features. But image interpretation is contextual, and users in specific situations look for images with specific objects or messages. While there are differences between the two parts, the semantic gap exists. Thus, to improve retrieval results, the semantic gap must be reduced.

The content-based image retrieval process can be broken into two steps: image description and image similarity measure. First, the image description step generates image signatures. Then, the image similarity measurement step uses image signatures to evaluate the distance between images. Image description is critical to image retrieval because it extracts the visual and semantic information of images, and good image description can help improve image retrieval results.

The image description step can be further broken into two steps: image feature extract and image signature construction. Image feature generation can be treated as a pre-process step for image signature construction. First, different types of image features are extracted globally or locally. Then, image signatures are constructed through different approaches mathematically or adaptively. Region-based image signature construction is a popular approach, which first segments images into segment regions and then constructs image signatures using these regions. Region-based approach treats each image segment as an image object, which is close to human vision. Then this approach uses the features of each image segment to build signatures for images. We will introduce more background knowledge to this step in Chapter 2.

After image signatures are constructed, we must determine how to use the signatures to measure the similarities between images. Similarity interprets the difference of an image with other images. Similarity techniques use "distance" to measure the dissimilarity between a pair of images. When retrieving images with a query image, the results are searched and ranked starting with smallest distance. Also, when retrieving images with the same semantic topics, images are categorized based on the closeness of distance.

In this thesis, we will introduce and evaluate a new similarity measure for image retrieval – potential energy distance. To evaluate the performance of potential energy distance, we conduct

experiments to assess the accuracy, speed, and robustness of the retrieved results, and we compare these results with two popular image measures – Mallows distance and IRM.

In the following sections of this Chapter, we will first summarize the features and approaches of similarity measures in Section 1.2. Then, we will review the literature on similarity measures in Section 1.3. Finally, we will identify the research questions of the thesis.

## 1.2. Similarity Measures

Different types of image features require different similarity measures. Figure 1 summarizes different types of measures and techniques for different image signatures. As we mentioned before, region-based signature is a popular and widely used approach; moreover, the related similarity measures and techniques have been extensively studied. In this thesis, we will use region-based signatures to construct image signatures using sets of vectors and associated weights.

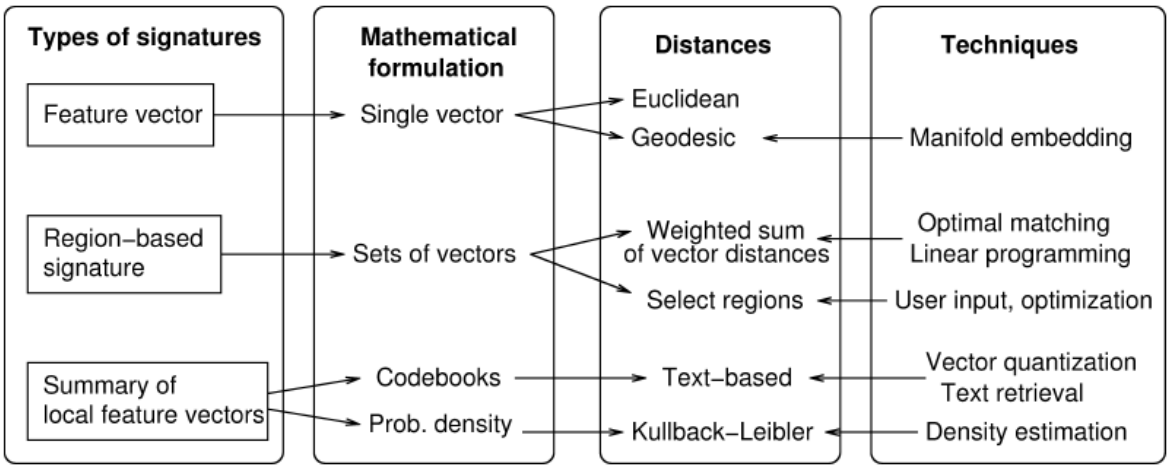


**Figure 1. Different Similarity Measures and Techniques for Different Types of Signatures [1]**

[1] summarized image similarity measures by their design philosophies. These techniques can be categorized based on the following features:

- Whether image features are represented as vectors, non-vectors, or ensembles;
- Whether images are compared globally or locally;
- Whether image similarity measures are in linear or nonlinear space;
- Whether image similarity measures are deterministic, stochastic, or fuzzy; and
- Whether image similarity measures use adaptive learning approaches.

During the past decades, various image similarity measures have been proposed. We will review these similarity measures in Section 1.3. To evaluate the performance of similarity measures, different evaluation measures have been proposed. These evaluation measures can be categorized from different perspectives:

- Precision. Precision evaluates the extent of coincidence of a similarity measure to image semantic content and human vision.
- Speed. Speed evaluates the computation efficiency of a similarity measure.
- Robustness. Robustness evaluates the extent of invariance of a similarity measure to image alternations.
- Object level comparison. Object-level comparison treats each image region as an object, and compares images based on image regions.
- Metric. If a measure is a metric, it means the distance between a pair of images follows triangle inequality, and the distance is zero only if the images are identical.

In this thesis, we introduce potential energy distance as a new image similarity measure. Potential energy distance was first proposed by Gabor J. Szekely in 1985. Potential energy

distance is a metric to measure the distance between probability distribution, which can be applied to measure image similarities. Potential energy distance is fast to compute because it assumes image probability distributions are independent and does not require optimal matching.

## 1.3. Related Research

Various similarity measurement techniques have been proposed. These measurement techniques can be roughly categorized into feature-based matching, region-based matching, and summarized local featured-based matching.

Feature-based matching was first proposed for image similarity measure. Feature-based matching techniques are used to compare image feature sets based on similarity functions. Using different selected features, these matching techniques can be categorized into different groups: Generic feature-matching approaches are used to compare an observed image feature set [3]. Object-silhouettes-based matching approaches are used to compare image shape features [4]. Structural feature-matching approaches are used to compare hierarchically ordered set of features [5]. Salient feature-matching approaches are used to compare interesting points in images [6]. Semantic level matching approaches are used to compare images using their semantic context [7]. Learning-based matching approaches are used to generate image semantics from learning experience instead of simply detecting visual features for different sematic terms [8, 9]. Learning-based matching approaches are available when the size of images becomes large. Different similarity functions are proposed to measure the distance of different image features. These feature-based matching approaches are efficient for computation, but they may be unable

to represent complex image semantics when images contain multiple image objects. In recent years, single-feature vector-based image retrieval approaches have been further studied to use nonlinear manifolds and geodesic distance [10-16]. It is argued that nonlinear subspace corresponds better with human perception.

Region-based matching approaches are proposed because single-feature vectors cannot represent complex image semantics. Region-based approaches segment images into sets of regions and treat each segmented region as an object in the original image. Region-based approaches then compare image similarities based on the similarities between sets of regions. Thus, region-based approaches depend greatly on the quality of image segmentation. In the ideal situation, each image object can be segmented into each region; in reality, however, accurate image segmentation is very challenging. Optimally selecting a subset of regions based on users' interests helps to avoid inaccurate segmentation. Probabilistic representation provides an alternative way to increase robustness and reduce the limitations of region-based approaches.

Summaries of local feature vectors are also used to represent images. Instead of representing an image by a set of segmented regions, summaries of local feature vectors use code book and probability density functions to generate image signatures. Codebooks are normally generated by vector quantization. Density functions can be generated by fitting a Gaussian mixture model [17], and Kullback-Leibler distance is used to measure the distance between distributions [18]. Approaches using summaries of local feature vectors are explored [18-23].

Besides the above mentioned approaches, other similarity measure approaches are proposed based on categorizing images or adaptive learning from user inputs [24-30]. Some probabilistic

frameworks are proposed to integrate image signature generation and image similarity measurement to minimize image retrieval errors [31-34].

Among these approaches, region-based approaches are widely used in the current decade. In this thesis, because we apply region-based matching techniques, we will mainly review research on region-based matching techniques in the following section. When comparing a pair of images, region-based matching approaches use different region-based similarity measures to compute the distance between image regions and then combine the distances between the regions to obtain the distance between images. While it is easy to define the distance between a pair of single vectors, it is not easy to define the distance for a set of vectors.

For image similarity measurement, various distances are used as distance functions. Euclidean distance is a widely used distance function. Euclidean distance is defined as:

$$D(X,Y) = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{m}(X_i - Y_i)^2} \quad (1)$$

where $X = (X_1, X_2, \ldots, X_n)$ and $Y = (Y_1, Y_2, \ldots, Y_m)$ are $n$-dimensional and m-dimensional vectors separately. More generally, Minkowski distance is defined to calculate n-norm distance, and it is defined as:

$$D(X,Y) = \sqrt[p]{\sum_{i=1}^{n}\sum_{j=1}^{m}(X_i - Y_i)^p} \quad (2)$$

Weighted Euclidean distance is also a widely used distance function. A weighted Euclidean distance-based matching approach assigns a weight to each pair of regions. The weight associated with each pair of regions measures the significance of correlating the pair of regions during the image-matching process. The distance between a pair of images is represented by the

aggregated weighted Euclidean distance between all pairs of regions. Weighted Euclidean distance is defined as:

$$D(X,Y) = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{m} w_{i,j}(X_i - Y_j)^2} \quad (3)$$

where $w_{i,j}$ is the weight between region $X_i$ and $Y_i$, and $w_i$ satisfies $0 < w_{i,j} < 1$ and $\sum_{i=1}^{n}\sum_{j=1}^{m} w_{i,j} = 1$.

Other distances commonly used in image retrieval include Canberra distance, Angular distance, Czekanowski coefficient[35], inner product, Dice coefficient, Cosine coefficient and Jaccard coefficient [36], Hausdoff distance, perceptually modified Hausdoff distance, weighted correlation distance [37], and quadratic form distance [38]. We give the definitions of these distances as follows, and $X = (X_1, X_2, \ldots, X_n)$ and $Y = (Y_1, Y_2, \ldots, Y_n)$ represent *n*-dimensional and m-dimensional vectors separately.

Canberra distance is applicable only to vectors with non-negative components, which suits color based vector similarity measurement. Canberra distance is defined as:

$$D(X,Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} \frac{|X_i - Y_j|}{|X_i| + |Y_j|} \quad (4)$$

Angular distance measures the size of the angle between the direction from the original point to one object and the direction from the original point to another object. Angular distance is meaningful to measure image similarity because in the color space similar colors have parallel orientations while different colors have different directions. Angular distance is defined as:

$$\theta(X,Y) = 1 - \frac{2}{\pi}\cos^{-1}\left(\frac{\vec{X_i}\cdot\vec{Y_j}}{|\vec{X_i}||\vec{Y_j}|}\right) \quad (5)$$

Czekanowski coefficient is also applicable only to vectors with non-negative components. Czekanowski coefficient is defined as:

$$D(X,Y) = \frac{2\sum_{i=1}^{n}\sum_{j=1}^{m}\min(X_i,Y_j)}{\sum_{i=1}^{n}\sum_{j=1}^{m}(X_i+Y_j)} \tag{6}$$

Inner product is defined as:

$$D(X,Y) = \sum_{i=1}^{n}\sum_{j=1}^{m}X_iY_j \tag{7}$$

Dice coefficient is defined as:

$$D(X,Y) = \frac{2\sum_{i=1}^{n}\sum_{j=1}^{m}X_iY_j}{\sum_{i=1}^{n}X_i^2+\sum_{j=1}^{m}Y_j^2} \tag{8}$$

Cosine coefficient is defined as:

$$D(X,Y) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}X_iY_j}{\sqrt{\sum_{i=1}^{n}X_i^2}\sqrt{\sum_{j=1}^{m}Y_j^2}} \tag{9}$$

Jaccard coefficient is defined as:

$$D(X,Y) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}X_iY_j}{\sum_{i=1}^{n}X_i^2+\sum_{j=1}^{m}Y_j^2-\sum_{i=1}^{n}\sum_{j=1}^{m}X_iY_j} \tag{10}$$

Hausdoff distance considers only the centroid structure of the vector features, while does not consider the weights of the vectors. Hausdoff distance is defined as:

$$D(X,Y) = \max\{\max_i \min_j d(X_i,Y_j), \max_j \min_i d(Y_j,X_i)\} \tag{11}$$

where $d(X_i,Y_j)$ is a ground distance function between $X_i$ and $Y_j$.

Perceptually modified Hausdoff distance is an extension of the Hausdoff distance, which considers both the centroid structure of the vector features and vector weights. It is defined as:

$$D(X,Y) = \max\{\frac{\sum_{i=1}^{n} w_i \, min_j\{\frac{d(X_i,Y_j)}{\min\{w_i,w_j\}}\}}{\sum_{i=1}^{n} w_i}, \frac{\sum_{j=1}^{m} w_j \, min_i\{\frac{d(Y_j,X_i)}{\min\{w_j,w_i\}}\}}{\sum_{j=1}^{m} w_j}\} \quad (12)$$

where $d(X_i, Y_j)$ is also a ground distance function between $X_i$ and $Y_j$.

Weighted correlation distance is defined as:

$$D(X,Y) = 1 - \sum_{i=1}^{n}\sum_{j=1}^{m} s(X_i,Y_j) \frac{w_i}{\sqrt{\sum_{i=1}^{n} w_i^2 s(X_i,X_i)^2}} \frac{w_j}{\sqrt{\sum_{j=1}^{m} w_j^2 s(X_j,X_j)^2}} \quad (13)$$

where

$$s(X_i,Y_j) = \begin{cases} 1 - \frac{3}{4}\frac{d(X_i,Y_j)}{R} + \frac{1}{16}\left(\frac{d(X_i,Y_j)}{R}\right)^3, & if\ 0 \le \frac{d}{R} \le 2 \\ 0, & otherwise \end{cases} \quad (14)$$

and $d(X_i, Y_j)$ is a ground distance function between $X_i$ and $Y_j$, and R is the maximum cluster radius.

Quadratic form distance is used to overcome the insensitivity of the Minkowski distance to other dimensions. It is defined as:

$$D(X,Y) = \sqrt{(X-Y)A\,(X-Y)^T} \quad (15)$$

where A is an n*n positive-definite matrix [37]. If matrix A is diagonal, then the quadratic form distance is the weighted Euclidean distance, and if matrix A is an identity matrix, then the quadratic form is the Euclidean distance.

Now we will introduce several basic matching techniques, including Mallows distance matching and IRM. These matching techniques vary in how they assign matching weights between regions, measure the distance between images based on distances between regions, and match one-one or many-many.

One-one match only allows one region to match one region across two images. In this matching mechanism, one region in the query image is matched with one best matched region in the target image [39]. After all regions are matched, the similarity of images is represented by the aggregate weighed similarity between all pairs of regions. Due to the challenge of image segmentation, each segmented region may not correspond to one image object. Thus, one-one match is not robust against inaccurate image segmentation. Many-many match is proposed to solve this issue. Many-many match allows one region to match several regions across two images. Mallows distance and IRM both belong to many-many match.

Mallows distance-matching method uses weighted Euclidean distance as its distance function. Mallows distance is a distance function for probability distributions in a metric space. The goal of Mallows distance is to find the minimum expected difference between the probability distributions of a pair of images optimized over all joint distributions [40]. Because the weight of a region indicates the significance of the region and the sum of matching weights of a region indicates the sum influence of this region to all other regions across two images, it is natural that these two quantities should be equal [40]. So the Mallows distance-matching approach tries to seek minimum image distance while subject to the relations between region weights and matching weights.

It is worth to note that the earth mover's distance (EMD) is a special case of Mallows distance [41]. EMD based approach is also a soft matching technique that uses sets of image regions. EMD treats image matching as transforming one distribution into another based on linear optimization while using minimum effort [42]. EMD is the special case of Mallows distance when the regional weights are probabilities.

IRM also uses aggregated weighted Euclidean distance to compute the distance between images, but unlike Mallows distance, IRM does not try to minimize the distance between images. Instead, it matches the pair of regions having the smallest distance with the highest priority and assigns a maximum valid weight as the matching weight between the pair of regions. This process is repeated until all regions are matched. Because of this, IRM is significantly faster than Mallows distance and performs no worse for image retrieval results [43].

Hausdoff distance is used to compute the distance between sets of regions. The Hausdoff distance-based approach uses the maximum distance among all matched regions as the distance between images [44]. The Hausdoff distance is then defined as the larger distance between the distance of a pair of images and the distance of this pair of images with reversed role.

In the recent years, enriched image retrieval approaches based on the above matching techniques were proposed. These approaches improved image retrieval results by generating image signatures, region weights, and image segments in more complex ways. Wang, et al. proposed to use an initial classification and a combination of different features for image retrieval [43]. Zhang and Zhang proposed to retrieve images based on the assumption that regions are generated probabilistically based on the hidden underlying semantic concepts [45]. Jing, el al. proposed an efficient approach by using vector quantization to build a region codebook from training images

[46]. Du and Wang first used feature-based vectors to categorize images and then used region-based vectors and IRM to narrow down the image retrieval scope [47].

As we mentioned before, accurate image segmentation is very challenging. Some approaches are proposed to reduce the dependency of image retrieval on image segmentation results. Chen and Wang proposed a variation of IRM, which used fuzziness against inaccurate image segmentation results [48]. Amores, et al. proposed the use of context and boosting techniques for image model learning and categorization which did not rely on the accuracy of image segmentation [49]. Also working against inaccurate image segmentation, Hoiem, et al. suggested a windowed search over location and scale [50]. Dagli and Huang proposed to use rectangular blocks to roughly segment foreground and background regions and then to retrieve images using only foreground regions [51].

Other region-based similarity measures are proposed for image retrieval. Mehul, et al. proposed a new similarity measure, complex wavelet structural similarity, which was robust to small image rotations and translations [52]. Beecks, et al. proposed the Signature Quadratic Form Distance, which bridged the gap between the traditional Quadratic Form Distance and image feature signatures [53-55]. Vazquez, et al. proposed to measure image similarity using the normalized compression that was first used to measure the similarity between strings [56].

Another approach against inaccurate segmentation is to select a subset of regions based on users' interests. When users search images, they may only be interested in a portion of the query image instead of the entire image. These situations require region-based queries. Region-based query approaches are proposed for region-based retrieval [44, 57, 58].

## 1.4. Research Questions

The research goal of this thesis is to introduce a new technique to measure image similarity. To evaluate the performance of this new measurement, we can compare it with the widely used benchmark measurements.

In the previous section, we listed several factors to evaluate the performance of similarity measures: accuracy, speed, robustness, object-level comparison, and metric. Szekely demonstrated that potential energy distance is a metric [59]. In this thesis, we apply potential energy distance with region-based matching techniques by measuring image similarity based on the distance between image regions; in other words, we compare images by using an object-level approach. To evaluate other performance aspects of potential energy distance, we propose the following research questions:

Q1. Does potential energy distance help improve the accuracy of image retrieval?

Q2. Does potential energy distance help improve the speed of image retrieval?

Q3. Is potential energy distance robust enough for image retrieval?

To answer these questions, we will implement experiments to evaluate the accuracy, speed, and robustness of potential energy distance. In the next section, chapter 2, we introduce background knowledge, including image segmentation techniques, region-based feature construction techniques, and image similarity measures, specifically Mallows distance and IRM. In chapter 3, we discuss details about potential energy distance. Our experiment, including data preparation, experiment design, and experiment results based on different evaluation approaches, is revealed in chapter 4, and we answer the research questions and conclude in chapter 5.

# Chapter 2 Background Knowledge

In this chapter, we provide background knowledge used in our approach. Image retrieval can be broken into two steps: image feature construction and image similarity measurement. Because image features serve as inputs, image feature construction can be treated as the preprocessing step to image similarity measurement. To retrieve images, we need to obtain image signatures based on the content and semantics of images first. In section 2.1, we will introduce approaches to generate image signatures, including image segmentation approaches and image feature construction approaches. In section 2.2, we introduce two widely used similarity measure techniques, Mallows distance and IRM.

## 2.1. Image Description

Image descriptions are critical to image retrieval because they can provide visual and semantic information even while the original images are in the format of pixel array, which poorly describes the content information of images. Image descriptions are used as input data for image similarity measurement, and good image descriptions can help improve image retrieval results.

Image description generation can be broken into two steps: image feature extract and image signature construction. Because image features are used as input for image signature construction, image feature extract can be treated as a pre-process step for image signature

construction. Image description generation workflow begins when sets of image features are extracted globally or locally, and then, image signatures are generated, using the image features mathematically or adaptively by active learning or user feedback.

Since 2000, various image features extract and image signature construction approaches have been proposed. The invention of these approaches directly promotes the development of new image similarity measurement techniques. In this thesis, we use region-based image signatures. In Section 2.1.1, we introduce image segmentation techniques, and in Section 2.1.2, we introduce region-based image signature construction approaches.

### 2.1.1. Image Segmentation

Based on how image features are extracted, we can roughly divide the current approaches to image segmentation into three categories: histogram extract, color-layout extract, and region-based extract [43]. Histogram approaches extract image features by image color histograms or distributions. While these approaches are fast and easy to compute, they are sensitive to image alternations. Color-layout approaches may be less sensitive to image alterations. They first partition images into blocks and then extract image features of each block. Unfortunately, such approaches sharply increase the computation complexity. To overcome the shortcomings of histogram and color-layout approaches, region-based approaches are proposed. Region-based approaches represent images at object level by treating each segmented region as an image object. [43] mentioned that region-based approaches are more close to human vision. Because they can identify objects in images, region-based approaches are also more robust to image alternations.

Region-based image features have received much interest over the past decade.

The first step to obtain region-based image features is to segment images into regions. Image segmentation techniques usually use clustering approaches to group image pixels. These techniques usually use image features as clustering feature vectors, such as color, shape, texture, and so on. In this way, pixels with similar features are grouped into the same cluster. Each cluster of pixels is subsequently treated as a segmented region [60]. The perfect result of image segmentation is shown when each region represents an object in the original image. In this way, we can get correspondences between segmented regions and physical objects in the image [61]. Image segmentation techniques emphasize the balance between segmentation results and computation complexity. Image segmentation techniques have been extensively studied and are commonly used in the fields of image analysis, including image comparison, image retrieval, and image semantic analysis [1, 62]. In this section, we introduce an image segmentation technique that we apply in our experiment, namely Multi-Stage Agglomerative Connectivity Constrained Clustering (MS-A3C).

***Multi-Stage Agglomerative Connectivity Constrained Clustering Approach***

The Multi-Stage Agglomerative Connectivity Constrained Clustering (MS-A3C) approach was proposed by Jia Li in 2011 [63]. The MS-A3C approach aims at solving the connectivity problem for image segmentation.

The generic requirement of clustering is to group similar objects into the same cluster while separating different objects into different clusters. For image segmentation, however, generic clustering approaches cannot guarantee that pixels in the same segment connect to each other

geometrically. To solve this problem, the MS-A3C approach was proposed to ensure that each segmented region is spatially connected. If the pixels in the segmented region are connected, then the region is connected.

MS-A3C approach first uses top-down k-means clustering approach to obtain over segmented images, and then uses a bottom-up agglomerative clustering approach to merge these small segments to obtain larger, connected segments. The top-down clustering step aims to gather initial small segments and identify the connectivity among the components. In this step, color feature is used to group pixels. The bottom-up agglomerative clustering step aims to collect connected segments by gradually merging the initial small segments, but only if they are connected. Visually similar segments are merged first, gradually using a distance combined of color, edge, and location features, and then similar segments are merged gradually, using a complex distance combined of color-based distance, edge-based distance, balanced partition measure, and Jaggedness measure.

The MS-A3C approach is proved to have higher accuracy and much faster speed in comparison with other state-of-the-art image segmentation approaches [63].

### 2.1.2. Feature Selection

Image features are used to describe the visual content of images. Currently, commonly adopted image features include color, texture, shape, and salient points [1]. When combined, these features can provide joint image signatures.

- Color. Color is a very active area for image features and is good for describing images recorded in frontal view. Color is normally represented in a three dimension color vector. LUV color is more widely used than RGB color because LUV color corresponds better with human vision. L is the luminance of the color, and U and V are the chromatic components.
- Texture. Texture is usually formed by wavelet coefficients in high-frequency bands. Texture, as a feature, is good at capturing the granularity and repetitive patterns of image surfaces.
- Shape. Shape is an efficient feature that is robust for image segments.
- Salient points. Salient points are effective for image affine transformations and illumination changes.

Color and texture features are more robust for digital images. Shape and salient points are useful for recognizing objects, but they are not very robust for digital images because shape and salient points are less robust to noisy backgrounds, changing angles, and occlusion. Thus, it is difficult to discover themes and match images. In this thesis, we only use the color feature.

Features can be extracted globally or locally. Global feature extraction obtains image features using the overall features of images, while local feature extraction obtains a set of image features for every pixel or small block using the average feature values across their neighbors. Global feature extraction approaches normally outperform local feature extraction approaches in speed, for both feature extraction and image similarity measure. But global feature extraction approaches may be sensitive to location and thus are too rigid to represent images [1]. In this thesis, we will use a global feature extraction approach.

After image features are extracted, we must determine how to use the features to construct image signatures. A natural way would be associate image features with a distribution. Various types of distributions are used to construct image signatures:

- Discrete distribution. Discrete distribution is commonly used for histograms and region-based feature vectors. An image with histograms and region-based signatures can be regarded as a set of vectors, and each vector is assigned with weight. Thus, an image can be represented by a set of feature vectors and corresponding weights assigned to the feature vectors. Normally, the weights of all regions sum to 1.
- Continuous distribution. Continuous distribution is more commonly used for local feature vectors because continuous distribution can provide more support vectors and more accurate results [18].
- Stochastic distribution. Stochastic distribution considers the spatial dependence of local feature vectors [64].

Although continuous distribution and stochastic distribution may be suitable for local features or other special kinds of features, the computation increases greatly. Since we use region-based feature vectors in this thesis, we will use discrete distribution to construct image signatures.

More formally, a region-based color feature vector *f* is represented as:

$$f = (\bar{l}, \bar{u}, \bar{v}) \tag{16}$$

$\bar{l}, \bar{u}, \bar{v}$ are the average L, U, V colors of all pixels in the region. The weight of a region $p$ is obtained by:

$$p = \frac{the\ size\ of\ all\ pixeds\ in\ the\ region}{the\ size\ of\ all\ pixels\ in\ the\ image} \tag{17}$$

Overall, image $I$ is represented as:

$$I = \{(f_1, p_1), (f_2, p_2), \ldots, (f_n, p_n)\} \quad (18)$$

$f_1, f_2, \ldots, f_n$ are feature vectors, $p_1, p_2, \ldots, p_n$ are associated weights, $n$ is the size of image regions.

Our approach uses two steps to construct image signatures. First, we segment images into regions. Second, we compute the average LUV color vectors and associated weights for each region. In this way, we can represent an image as a vector of color components along with the distribution of the segmented regions.

## 2.2. Similarity Measures

After we obtain image signatures, the next step is to measure the similarity between a pair of images. Distance is used to represent the dissimilarity between images. The distance used to measure the dissimilarity should be consistent with human visual perceptions. Defining the distance between a pair of images is equivalent to defining the distance between sets of vectors because every feature vector in the vector set corresponds to an image region. Although the distance between two vectors is easily defined, for example Euclidean distance, the distance between sets of vectors is not easily defined [1].

The distance between a pair of images is normally obtained either by directly applying distance functions or by optimizing distance functions. Various distance functions have been proposed to measure the image distance. In section 1.3, we review related research on various distance

functions and similarity measures. In this section, we introduce two widely used benchmark similarity measures, Mallows Distance and IRM.

### 2.2.1. Mallows Distance

Mallows distance was first introduced to measure image distance by Mallows in 1972 [40]. Mallow distance is obtained by matching each region of an image with each region of another image, as shown in Figure 2. The goal of Mallows distance is to discover the minimum expected difference between the feature vectors of images. To achieve this goal, the key step of Mallows matching is to seek optimal matching weights between regions. Mallows distance applies linear programming to solve this problem.

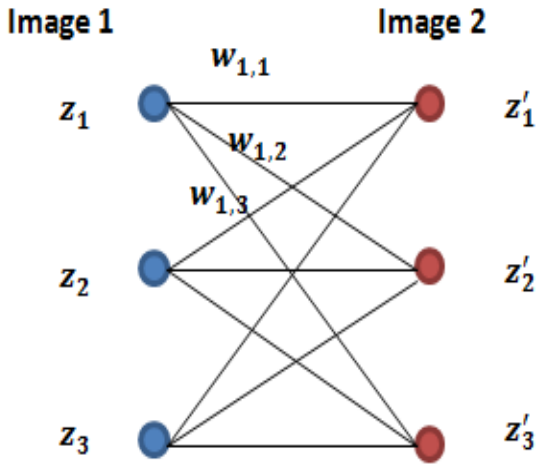


**Figure 2. Mallows Distance Region Matching [65]**

The distance between two regions is calculated using n-dimension Euclidean distance. The distance between a pair of images is represented as:

$$D(I_1, I_2) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} w_{i,j}\, d(f_i, f_j) \tag{19}$$

$w_{i,j}$ is the matching weight between region $i$ in image $I_1$ and region $j$ in image $I_2$, $d(f_i, f_j)$ is the Euclidean distance between feature vector of region $i$ and feature vector of region $j$.

The goal of Mallows distance is to get minimized expectation difference by assigning appropriate matching weighting $w_{i,j}$ between each pair of regions:

$$D(I_1, I_2) = min_{w_{i,j}} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} w_{i,j}\, d(f_i, f_j) \tag{20}$$

The constraints of Mallows matching are obtained using the relations between region weights and matching weights, the sum of matching weights between one region in an image and all regions in another image is equal to the weight of this region. This measurement is natural because we use region weight to measure the importance of a region and the sum of matching weights of a region to measure the total influence of this region to all its matching regions. The weight of a region equals to the sum of all matching weights of the region:

$$\sum_{j=1}^{n_2} w_{i,j} = p_i \tag{21}$$

$$\sum_{i=1}^{n_1} w_{i,j} = p_j \tag{22}$$

$w_{i,j}$ is the matching weight between region $i$ and region $j$, $p_i$ and $p_j$ are the weights of region $i$ and region $j$.

Also, as we mentioned in section 2.1.2, the weight of region is the percentage of the size over the image, so the sum of all region weights is equal to 1.

$$\sum_{i=1}^{n_1} p_i = 1 \tag{23}$$

$$\sum_{j=1}^{n_2} p_j = 1 \tag{24}$$

Also, basic requirement of weight needs to be satisfied:

$$w_{i,j} \geq 0 \tag{25}$$

Mallows distance is proved to be a metric [40], which can be applied for image similarity measure. The computation of Mallows distance is usually costly because it matches all regions between a pair of images. Computing Mallows distance is the same as the minimum cost flow problem, and the complexity is $O(N^3 \log N)$ [66].

### 2.2.2. Integrated Region Matching

As we introduced in previous sections, region-based image retrieval approaches segment images into sets of regions and treat each region as an image object. After image objects are extracted, region-based approaches compute the distance between images based on the distance of sets of image objects. In the ideal situation, each segmented region represents one semantic object in the original image, but in reality, ideal results are difficult to obtain. One reason is objects in images are 2D, but objects in reality are 3D [2].

Many region-based similarity measures compare similarities between images by comparing individual image regions [67, 68]. The shortcoming of such approaches results from the difficulty of image segmentation, specifically that a segmented region may not represent a single object. So the approaches that compare individual regions may not tolerate poor image

segmentation results. On the other hand, it is argued that if a similarity measure takes all image regions into consideration during image comparison, then it is more robust to image segmentation results [43].

IRM was proposed by Wang, et al. in 2000. IRM takes all image regions into consideration to calculate the overall similarity between images by applying two matching principles. First, IRM allows one-to-many region matching, which addresses the issue that an image object may be segmented into multiple regions due to inaccurate image segmentation. Second, IRM matches the most similar pair of regions first and then assigns the maximum valid weight to this pair of regions.

Like Mallows distance, IRM also uses Euclidean distance, but the goal of IRM is not to minimize the overall difference by assigning appropriate matching weighting between image regions. Instead, IRM keeps looking for the most similar pair of regions and matches this pair of regions first with the maximum valid weight. The maximum valid weight is the smaller available weight between the pair of the matching regions. After regions are matched, the maximum valid weight is subtracted from the available weights of the matching regions. If there is no weight available for a region, then this region will be considered fully matched. Otherwise, the available weight is used for the following matching. This process is repeated until all region weights are consumed. At last, the overall similarity between a pair of images is calculated by the weighted sum of similarities between each pair of regions. The process of IRM is illustrated in Figure 3. Note that Figure 3 shows that the IRM allows one region in an image to be matched with several regions in another image.

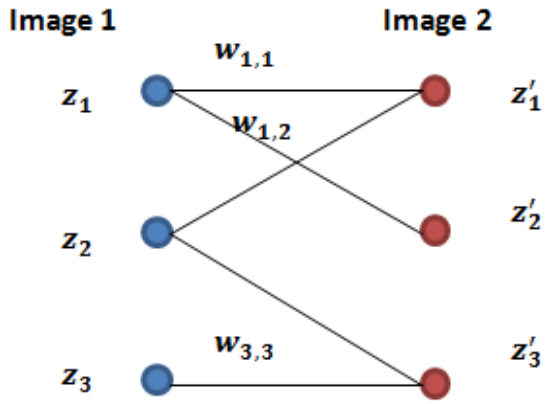


**Figure 3. IRM Matching Process [43]**

More formally, the distance between a pair of images is represented as:

$$D(I_1, I_2) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} w_{i,j}\, d(f_i, f_j) \tag{26}$$

$w_{i,j}$ is the matching weight between region $i$ in image $I_1$ and region $j$ in image $I_2$, $d(f_i, f_j)$ is the Euclidean distance between feature vector of region $i$ and feature vector of region $j$.

First, the distances between each pair of regions are computed. IRM applies "most similar highest priority" principle and subtracts maximum valid weight each time after region matching. The maximum valid weighted is defined as the smaller available weight between the two matching regions:

$$w_{i,j} = \min(p_i, p_j) \tag{27}$$

$p_i$, $p_j$ are the available weights of region $i$ and region $j$. Without loss of generality, assume $p_i \leq p_j$, so after each match the available weight of region $i$ is 0, and the available weight of region $j$ is $p_j - w_{i,j}$. In summary, after each match we have:

$$\sum_{j=1}^{n_2} w_{i,j} = p_i, \quad i \neq i' \tag{28}$$

$$\sum_{j=1}^{n_2} w_{i,j} = 0, \quad i = i' \tag{29}$$

$$\sum_{i:1\leq i\leq m, i\neq i'} w_{i,j} = p_j, \quad j \neq j' \tag{30}$$

$$\sum_{i:1\leq i\leq m, i\neq i'} w_{i,j} = p_{j'} - p_{i'}, \quad j = j' \tag{31}$$

Also, basic requirement of weight needs to be satisfied:

$$\sum_{j=1}^{n_2} w_{i,j} = p_i, \sum_{i=1}^{n_1} w_{i,j} = p_j \tag{32) (33}$$

$$\sum_{i=1}^{n_1} p_i = 1, \sum_{j=1}^{n_2} p_j = 1 \tag{34) (35}$$

$$w_{i,j} \geq 0 \tag{36}$$

After each region match, IRM excludes the region *i* from the matching process because region *i* has no available weight and is considered fully matched. Region *j* will be kept for further matching if it still has available weight. In this way, IRM enables one region in an image to match multiple regions in another image. This process is recursive until the weights of all regions are assigned.

IRM is significantly faster to compute than Mallows distance because IRM does not optimize matching weights to achieve minimum overall image distance as the Mallows approach does. The complexity of IRM is $O(N^2)$. Furthermore, Wang, et al. proved that IRM performed as well as, if not better than, Mallows distance in retrieval accuracy. Significantly, IRM is not a metric.

# Chapter 3 Potential Energy Distance

Newton's potential energy theory illustrates that the energy of an object or a system is generated by the position of the body or the arrangement of the components of the system [59]. Inspired by the potential energy theory, Szekely proposed potential energy distance to measure the distance between statistical probability distributions in 1985. Potential energy distance assumes that statistical components are heavenly bodies under statistical potential energy, and potential energy distance is generated between these statistical observations. Potential energy distance is used as a new measure for multivariate distribution based on Euclidean distance between statistical observations.

Potential energy distance has been applied to various areas of statistics, such as hierarchical clustering, testing multivariate normality, testing the multi-sample hypothesis of equal distributions, change point detection, multivariate independence including distance correlation and Brownian covariance, and scoring rules [59].

Potential energy distance is defined as

$$D^d(X, Y) = 2E||X - Y||^d - E||X - X'||^d - E||Y - Y'||^d \quad (37)$$

where $X = (X_1, X_2, ..., X_m)$ and $Y = (Y_1, Y_2, ..., Y_n)$ are $d$-dimensional distributions, $X'$ is an independent and identical copy of $X$, and $Y'$ is an independent and identical copy of $Y$, $\|X\| = (X^T X)^{1/2}$ is the Euclidean distance of X, $d \geq 1$.

Potential energy is proved by Szekely that if the following conditions are satisfied:

(1) $X$ and $Y$ are independent random vectors

(2) $X$ and $X'$ are independently and identically distributed

(3) $Y$ and $Y'$ are also independently and identically distributed

(4) d is a constant which make $E||X||^d$ and $E||Y||^d$ are finite

Then the energy distance between distributions of $X$ and $Y$ satisfies:

$$D^d(X,Y) = 2E||X-Y||^d - E\left||X - X'|\right|^d - E\left||Y - Y'|\right|^d \geq 0 \tag{38}$$

Szekely focused on two special cases: $d = 1$ and $d = 2$. He also mentioned that the result using $d = 1$ out-performs $d = 2$ when the distributions are different and the means of clusters are the same or close to each other. Also, when $0 < d < 2$, the distance between two statistical observations is equal to zero if and only if the two observations are the same in distribution. This means when $0 < d < 2$, potential energy distance is a metric. But when $d \geq 2$, the zero distance does not hold when $X$ and $Y$ have the same distribution.

In this thesis, we apply potential energy distance with $d = 1$:

$$\mathrm{D(X,Y)} = \sqrt{(2E\left||X-Y|\right| - E\left||X-X'|\right| - E\left||Y-Y'|\right|)} \tag{39}$$

where $E\left||X-Y|\right|$ is the expected distance between X and Y,

$$E\left||X-Y|\right| = w_{i,j}d_{i,j} \tag{40}$$

$w_{i,j}$ is the matching weight of region i and region j, $d_{i,j}$ is the Euclidean distance between region I and region j. Because $X$ and $Y$, $X$ and $X'$, and $Y$ and $Y'$ are all independently distributed, the matching weight between $X$ and $Y$, $w_{i,j}$ satisfies:

$$w_{i,j} = p_i p_j \tag{41}$$

where $p_i$, $p_j$ are the weights of region $i$ and region $j$. Thus, Equation (18) is equivalent to

$$D\,(X,Y)\; = \sqrt{\left(2\sum_{i=1}^{m}\sum_{j=1}^{n} p_i p_j\, d_{i,j} - \sum_{i=1}^{m}\sum_{i'=1}^{m'} p_i p_{i'}\, d_{i,i'} - \sum_{j=1}^{n}\sum_{j'=1}^{n'} p_j p_{j'}\, d_{j,j'}\right)} \tag{42}$$

where *m, n, m', n'* are the number of segmented regions of images *X, Y, X', Y'*. The complexity of the potential energy distance approach is $O(N^2)$.

In this thesis, we use the distance in equation (19) to measure the similarity between images. Because potential energy distance does not require optimization of matching weights to achieve minimum overall distance between images, which Mallows distance requires, this approach should offer computation efficiency. To evaluate the computation performance of potential energy distance, we implement an experiment in the next chapter to measure the speed of this approach, and we compare potential energy distance with Mallows distance and IRM. Additionally, we evaluate the performance of potential energy distance from the aspects of accuracy and robustness.

# Chapter 4 Experiments

In this chapter, we evaluate the performance of potential energy distance in image retrieval from different aspects and compare potential energy distance with two benchmark approaches: Mallows distance measure and IRM. In Section 4.1, we introduce the dataset used in our experiment and explain how we pre-processed the dataset to generate image signatures. Different measures for performance evaluation are presented in Section 4.2. Finally in Sections 4.3 – 4.5, we evaluate and compare image retrieval results from the perspectives of accuracy, speed, and robustness, before concluding the chapter.

## 4.1. Data Preparation

### 4.1.1. Datasets

In our experiment, we use the MIR Flickr benchmark image dataset, which was built by Mark J. Huiskes and Michael S. Lew in 2008. This dataset contains 25, 000 images downloaded from the Flickr photo sharing website. All images in this dataset are grouped into different categories. Also, in the MIR Flickr dataset, the tag and EXIF information associated with images are available for processing.

The 25,000 images in the MIR Flickr dataset are uploaded by thousands of Flickr users over a period of approximately 15 months. These images are representative of a generic domain,

including people, animals, outdoor scenes, portraits, autos, structures, and so on. The MIR Flickr dataset is guaranteed to include only images with high “interestingness”, as this dataset includes the 500 most interesting images from each day over the 15 month period [69]. The MIR Flickr dataset uses image scores to measure the quality of images in order to guarantee the “interestingness” of images. Image scores include factors such as where the clickthroughs on the image are from, who makes comments and when, and who marks the image as a favorite.

The topics of the MIR Flickr dataset are generated in two dimensions:

- Relevance level. The topics in this dataset are built gradually from possibly relevant level to actually relevant level in two steps. First, images are interpreted with topics in a very wide sense. The topics assigned to each image should capture the content of the image for various potential image searches. These topics are called potential labels, and they can act as a common topic collection that makes subsequent annotation of narrow interpretations faster. Second, images in this dataset are categorized using only relevant topics from the potential topics if the annotators think the images are really relevant to the topics.
- Abstraction level. The topics in this dataset are built from general topic level to specific topic level in two steps. The image collection is annotated initially using general topics that are selected by picking a couple of topics that can cover the topics of the images in this dataset. Subtopics are then assigned to the images that have a potential label for the more general topics. Table1 lists the general topics and subtopics used in the MIR Flickr dataset.

**Table 1. General Topics and Subtopics in MIR Flickr Dataset**

| General Topic | Subtopics |
|---|---|
| sky | clouds |
| water | sea/ocean, river, lake |
| people | portrait, boy/man, girl/woman, baby |
| night | |
| plant life | tree, flower |
| animals | dog, bird |
| man-built structures | architecture, building, house, city/urban, bridge, road/street |
| sunset | |
| indoor | |
| transports | car |

The general topics and subtopics are then grouped jointly into 24 categories. Table 2 lists all categories and the image size of each category.

**Table 2. Image Categories in MIR Flickr Dataset**

| ID | Categories | Size of Category |
|---|---|---|
| 1 | Animals | 3216 |
| 2 | Baby | 259 |
| 3 | Bird | 742 |
| 4 | Car | 1177 |
| 5 | Clouds | 3700 |
| 6 | Dog | 684 |
| 7 | Female | 6184 |

| | | |
|---|---|---|
| 8 | Flower | 1823 |
| 9 | Food | 990 |
| 10 | Indoor | 8313 |
| 11 | Lake | 791 |
| 12 | Male | 6081 |
| 13 | Night | 2711 |
| 14 | People | 10373 |
| 15 | Plant Life | 8763 |
| 16 | Portrait | 3931 |
| 17 | River | 894 |
| 18 | Sea | 1322 |
| 19 | Sky | 7912 |
| 20 | Structures | 9992 |
| 21 | Sunset | 2135 |
| 22 | Transport | 2850 |
| 23 | Tree | 4683 |
| 24 | Water | 3331 |

### 4.1.2. Image Signature Construction

In this section, we introduce how we pre-process the image dataset to generate image signatures. The workflow of the data pre-processing is illustrated in Figure 4.

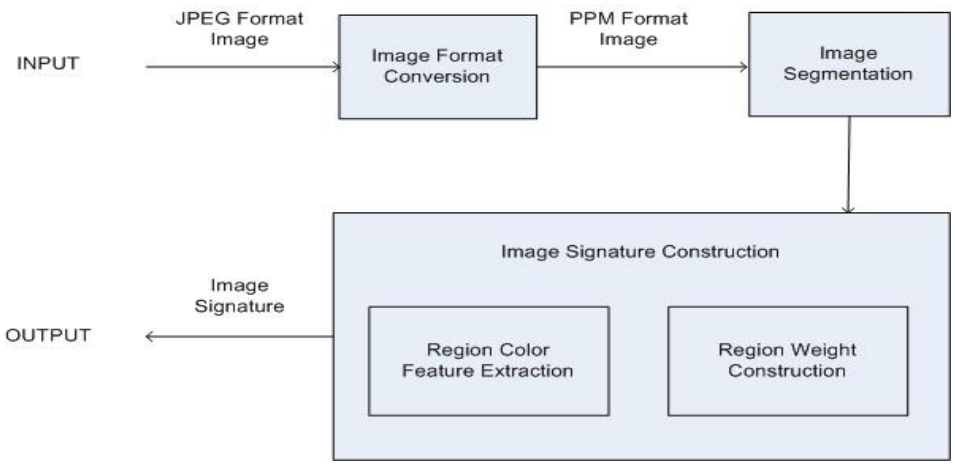


**Figure 4. Workflow of Image Signature Construction**

The steps of image signature construction workflow are as follows:

Step 1: Image format conversion. The images included in the MIR Flickr dataset use JPEG format. In this step, we convert the JPEG images to PPM format for ease of writing and analysis. PPM format contains little information besides basic color, but because we only use color features to generate image signatures in our experiment, the loss of other features due to image format conversion does not affect our experiment.

Step 2: Image segmentation. In our experiment, we use region-based features to construct image signatures. In this step, we use MS-A3C approach, as introduced in Section 2.1.1, to perform image segmentation. Ee use the open-source package provided by Li [63]. This package is implemented using Linux C language.

Step 3: Image signature construction. After we obtain segments of images, we generate image signatures using the color feature vectors of sets of segmented regions. The color feature vector of a region is obtained by averaging the LUV colors of all pixels in this region. The weight of a

region is obtained by the coverage of this region over the whole image. Image feature vector and associated weight of a region are represented as:

$$F = \begin{pmatrix} f_{1,L} & f_{1,U} & f_{1,V} \\ f_{2,L} & f_{2,U} & f_{2,V} \\ \cdots & \cdots & \cdots \\ f_{m,L} & f_{m,U} & f_{m,V} \end{pmatrix}, P = \begin{pmatrix} p_1 \\ p_2 \\ \cdots \\ p_m \end{pmatrix} \quad (43)$$

$F$ is the feature vector of an image, $m$ is the number of image segments, $f_{i,L}$, $f_{i,U}$, $f_{i,V}$ are the LUV features of region $i$, $P$ is the weight vector of the image, and $p_i$ is the weight of region $i$.

## 4.2. Evaluation Measures

To evaluate the performance of image retrieval approaches, first, we need to define which retrieval images are considered to be a "match". Usually, images are considered to be a match if they belong to the same categorization or their annotations are similar, which means they have the same themes. In our experiment, we consider images a match if they belong to the same category. Because the MIR Flickr dataset has already been categorized into 24 categories, we use these categories to evaluate whether images are matched.

As mentioned previously, we evaluate potential energy measure through accuracy, speed, and robustness aspects. In this thesis, we use precision and weighted precision to evaluate image retrieval accuracy, CPU time to evaluate image retrieval speed, and rank of the original image and potential energy distance between the altered image and the original image to evaluate image retrieval robustness.

Precision is a popular measure for performance evaluation. Precision can be computed using the confusion matrix proposed by Kohavi and Provost in 1998 [70]. The confusion matrix is a 2-by-2 matrix, and its elements are information about actual and predicted classifications. Table 3 displays the structure and elements of a confusion matrix:

**Table 3. Evaluation Confusion Matrix**

| Confusion Matrix | | Correct result | |
|---|---|---|---|
| | | E1 | E2 |
| Obtained result | E1 | tp (true positive) | fp (false positive) |
| | E2 | fn (false negative) | tn (true negative) |

The elements in the confusion matrix are in four categories: tp (true negative) is the number of positive objects correctly labeled, fp (false positive) is the number of negative objects incorrectly labeled, fn (false negative) is the number of positive objects incorrectly labeled, and tn (true negative) is the number of negative objects correctly labeled.

Precision is the proportion of the number of positive objects correctly labeled (tp) to the number of all objects labeled as positive (tp + fp). A high precision means that objects labeled as positive are highly relevant. Precision is defined as:

$$precision = \frac{\mathrm{tp}}{\mathrm{tp+fp}} \tag{44}$$

Weighted precision takes into account the rank of matched images. Earlier retrieved images are considered to be more similar to the testing image because they have less distance to the query image. Weighted precision is the weighted percentage of matched images. It assigns larger

weights to earlier retrieved images and assigns smaller weights to the later retrieved images. Weighted precision is defined as:

$$weighted\ precision = \frac{1}{n}\sum_{k=1}^{n}\frac{tp_k}{tp_k + fp_k} \tag{45}$$

where k is the retrieved number of images, $tp_k$ is the number of positive objects correctly labeled in the first k images, $fp_k$ is the number of negative objects incorrectly labeled in the first k images.

## 4.3. Accuracy

In this section, we evaluate the accuracy of potential energy distance as well as the accuracy performance of both image queries and image categorization.

### 4.3.1. Image Queries

In the image query experiment, we compare potential energy distance with two popular image retrieval measures, Mallows distance and IRM. To provide numeric results, we randomly select 480 sample images from all 24 categories, each category containing 20 images. Image retrieval is performed on all 25,000 images in the MIR Flickr dataset. A retrieved image is considered a match if and only if it is in the same category with the query image. Also, because precision and weighted precision will vary depending on the number of retrieved images, we compute precision and weighted precision for the top 10, top 20, …, top 100 retrieved images separately.

Due to the limit of space, in this section we only display and analyze results for the top 10, top 50, and top 100 retrieved images in each category. In the section of image categorization, we display and analyze experiment results for the top 10, top 20, …, top 100 retrieved images in each category because we use all images in the MIR Flickr dataset as query samples in that experiment. Figures 5 – 7 list the average precisions and weighted precisions of top 10, top 50, and top 100 retrieved images of each category.

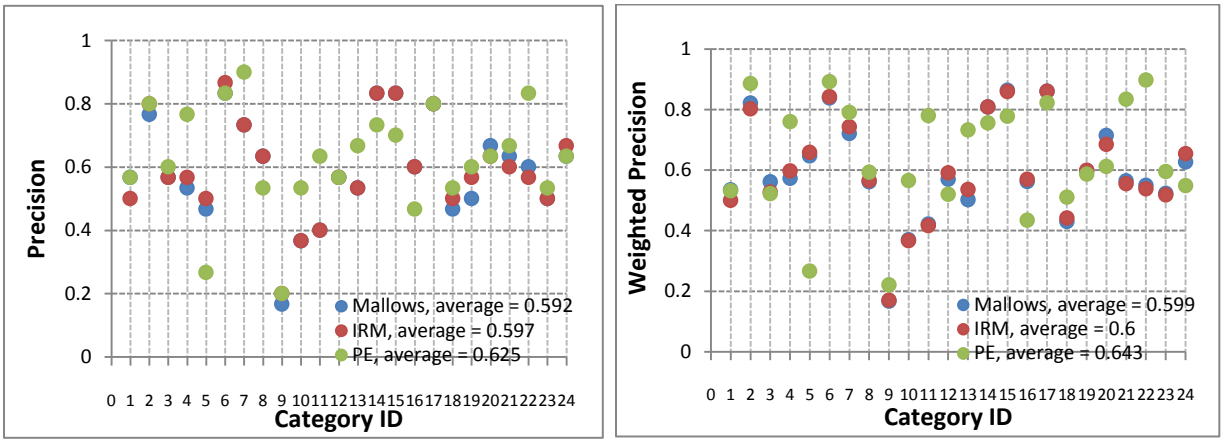


**Figure 5. Average and Weighted Precisions of Top 10 Retrieved Images in Image Query**

As shown in Figure 5, we see that overall potential energy distance performs slightly better than Mallows distance and IRM. The average precision of potential energy distance is higher than Mallows distance by 0.033 and higher than IRM by 0.028. The weighted precision of potential energy distance is higher than Mallows distance by 0.044 and higher than IRM by 0.043.

We can also see that among all image categories, potential energy distance performs significantly better than do the other two approaches in both average precision and weighted precision for

categories "Car," "Indoor," "Lake," "Night," and "Transport," and performs significantly better in weighted precision for category "Sunset." However, Mallows distance and IRM perform significantly better than potential energy distance in average precisions and weighted precisions for categories "Clouds," "People," "Plant Life," and "Portrait."

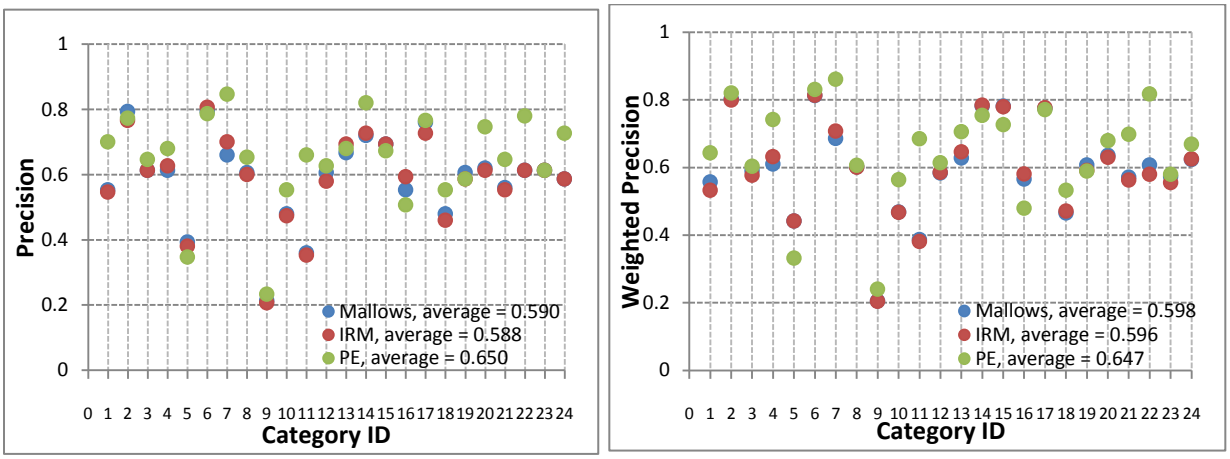


**Figure 6. Average and Weighted Precisions of Top 50 Retrieved Images in Image Query**

From Figure 6, we can see that overall potential energy distance performs slightly better than Mallows distance and IRM. The average precision of potential energy distance is higher than Mallows distance by 0.06 and higher than IRM by 0.062. The weighted precision of potential energy distance is higher than Mallows matching by 0.049 and higher than IRM by 0.051.

Among all 24 image categories, potential energy distance performs significantly better than the other two approaches in average precision and weighted precision for categories "Lake" and "Transport."

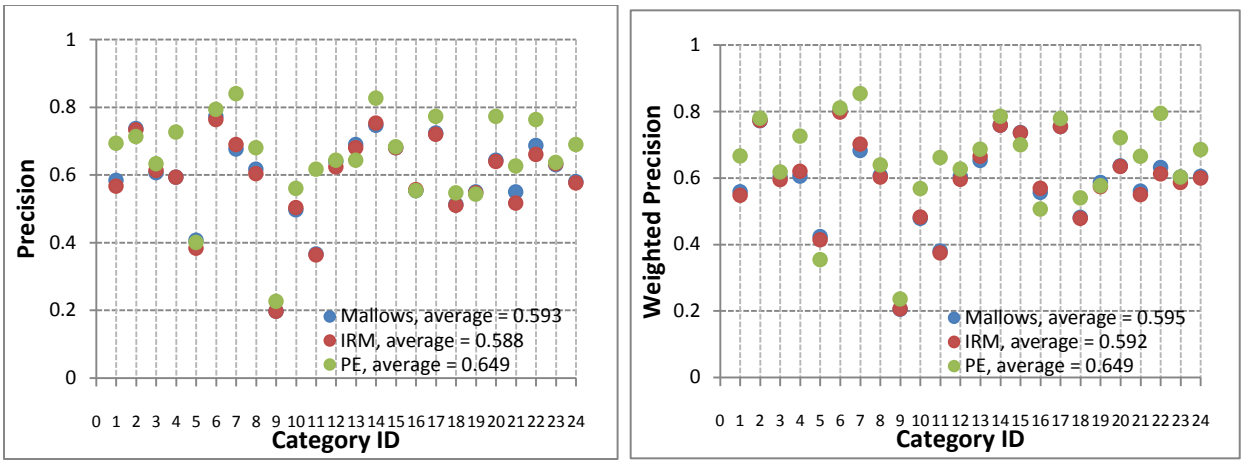


**Figure 7. Average and Weighted Precisions of Top 100 Retrieved Images in Image Query**

Figure 7 reveals that overall potential energy distance performs slightly better than Mallows distance and IRM. The average precision of potential energy distance is higher than Mallows distance by 0.056 and higher than IRM by 0.061. The weighted precision of potential energy distance is higher than Mallows distance by 0.054 and higher than IRM by 0.057.

We can also see that among all 24 image categories, potential energy distance performs significantly better than the other two approaches in average precision and weighted precision for category "Lake."

Also, from Figures 5 - 7, we can see that average precision and weighted precision in each category vary a little with different numbers of retrieved images. We will analyze this result in more detail in the next section when we analyze all top 10, top 20, …, top 100 retrieved results.

### 4.3.2. Image Categorization

In the image categorization experiment, each image in the MIR Flickr dataset is used as a query image, and image retrieval is performed over all images in this dataset. We still consider a retrieved image a match if and only if it is in the same category with the query image. Similar to Section 4.3.1, we compute precision and weighted precision for the top 10, top 20, …, top 100 retrieved images separately because precision and weighted precision will vary depending on the number of retrieved images. Figures 8 – 17 list the average precisions and weighted precisions of top 10, top 20, ..., top 100 retrieved images in each category.

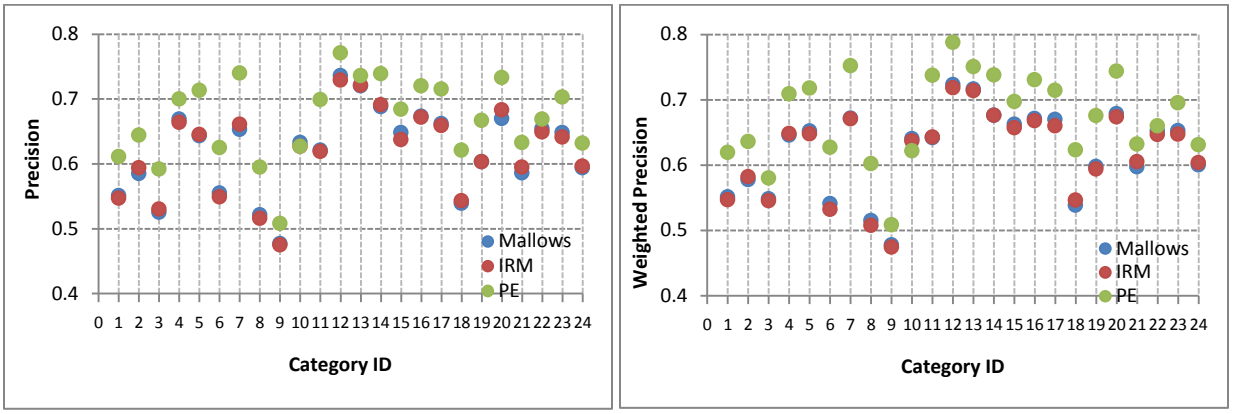


**Figure 8. Average and Weighted Precisions of Top 10 Retrieved Images in Image Categorization**

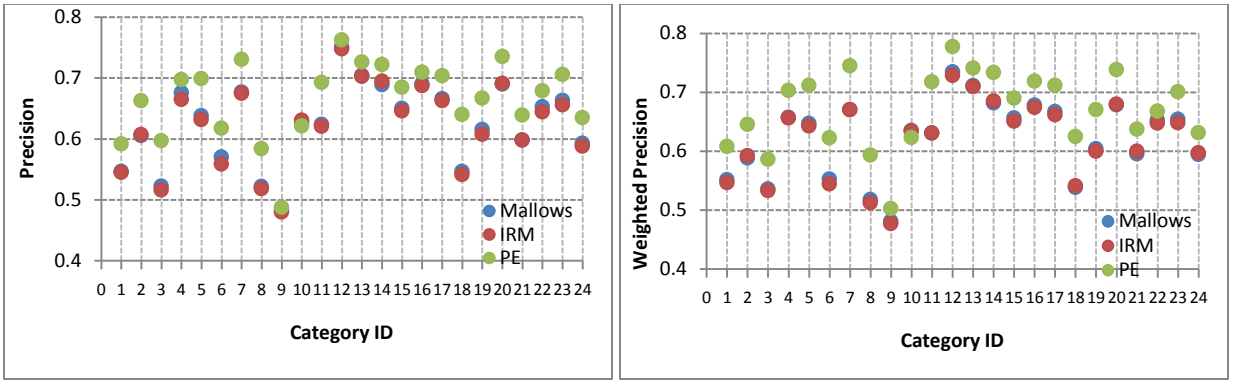


**Figure 9. Average and Weighted Precisions of Top 20 Retrieved Images in Image Categorization**

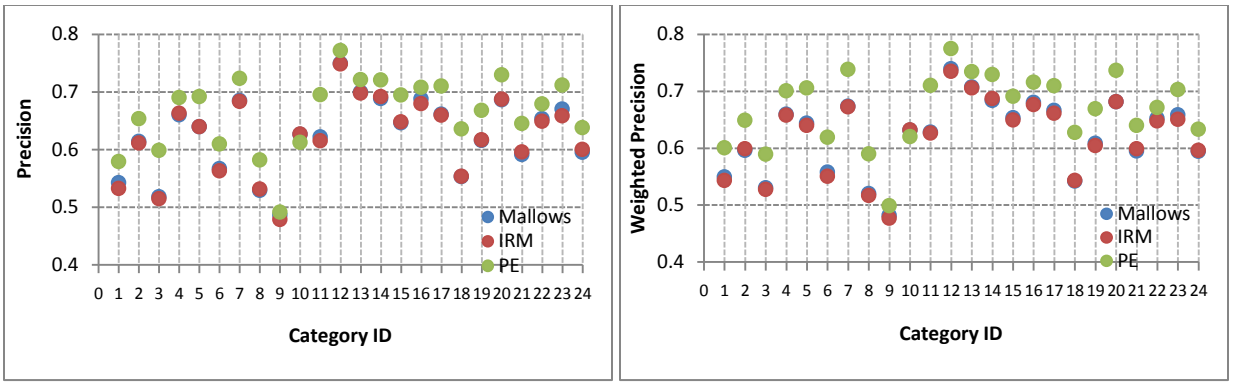


**Figure 10. Average and Weighted Precisions of Top 30 Retrieved Images in Image Categorization**

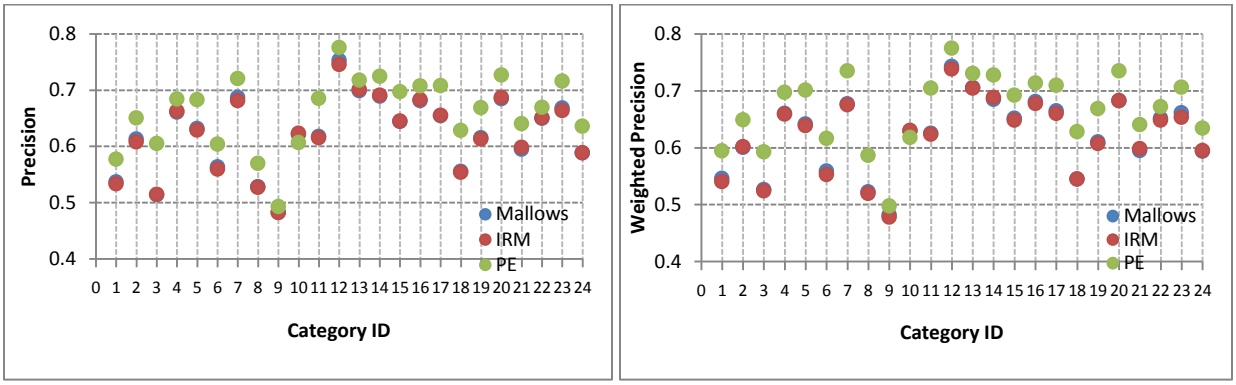


**Figure 11. Average and Weighted Precisions of Top 40 Retrieved Images in Image Categorization**

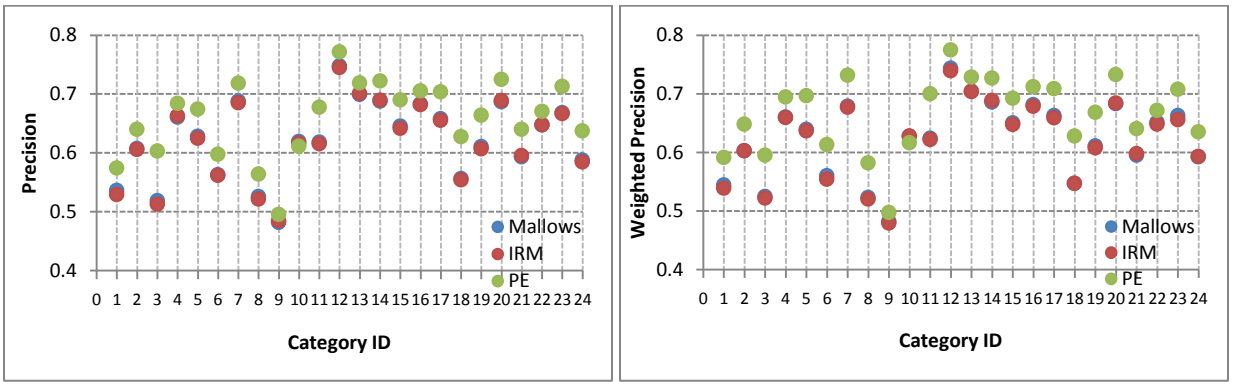


**Figure 12. Average and Weighted Precisions of Top 50 Retrieved Images in Image Categorization**

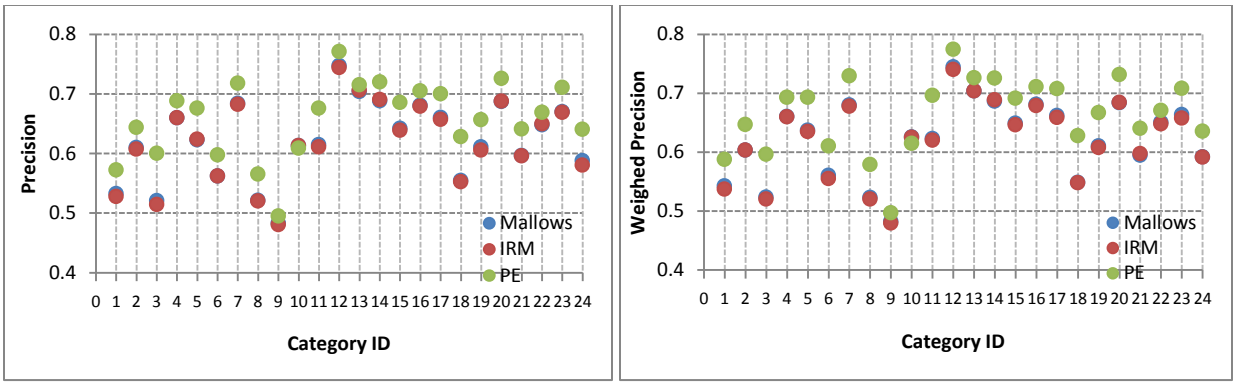


**Figure 13. Average and Weighted Precisions of Top 60 Retrieved Images in Image Categorization**

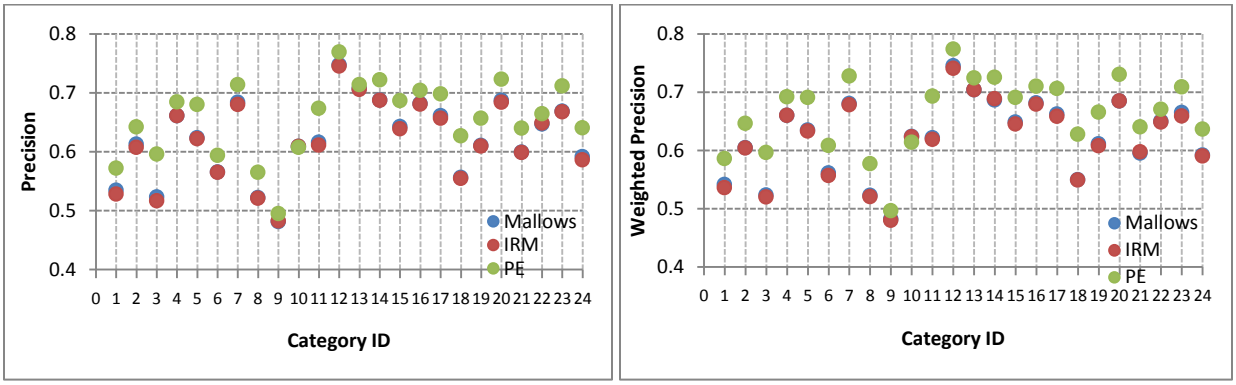


**Figure 14. Average and Weighted Precisions of Top 70 Retrieved Images in Image Categorization**

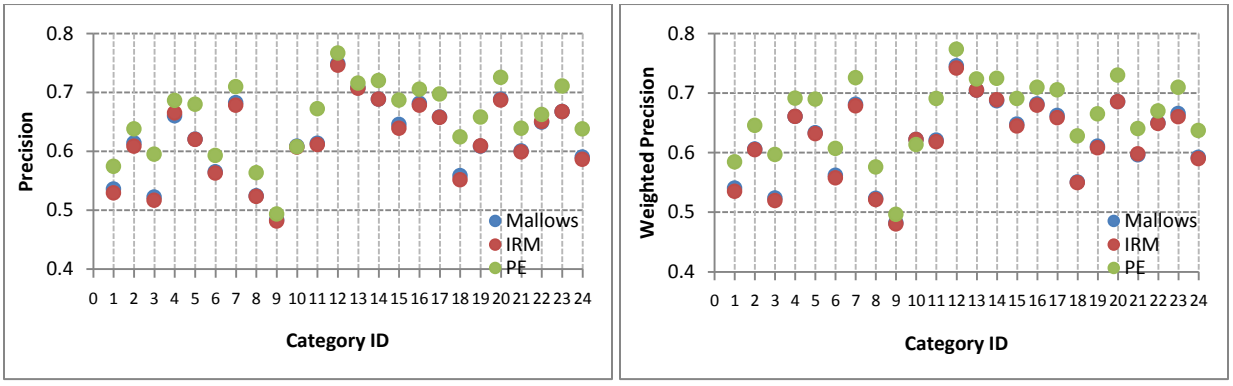


**Figure 15. Average and Weighted Precisions of Top 80 Retrieved Images in Image Categorization**

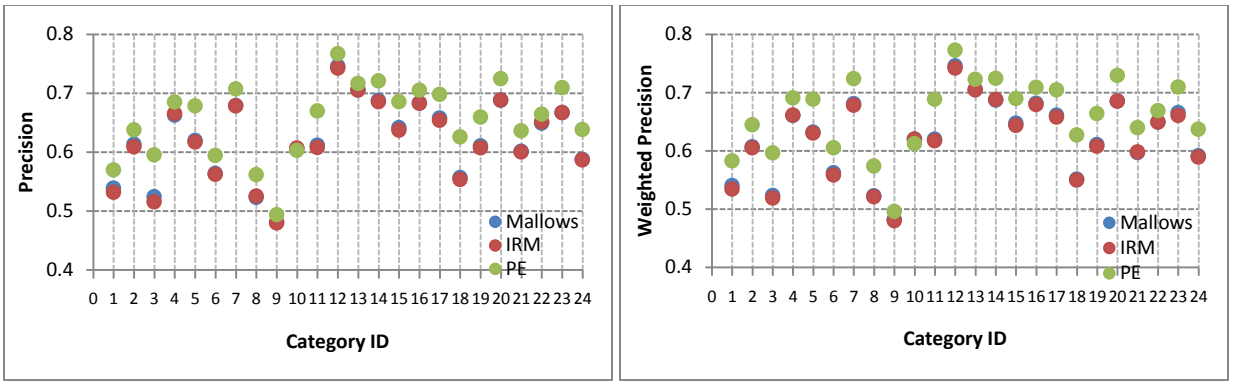


**Figure 16. Average and Weighted Precisions of Top 90 Retrieved Images in Image Categorization**

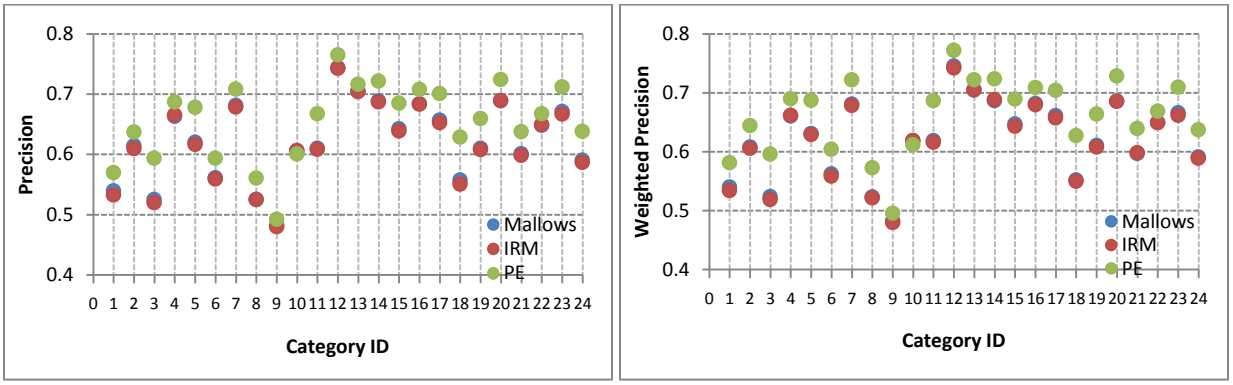


**Figure 17. Average and Weighted Precisions of Top 100 Retrieved Images in Image Categorization**

For this experiment, average precision and weighted precision does not vary much with different numbers of retrieved images, as shown in Figures 8-17. We will take the category "Animals" for example. Figure 18 shows the trends of average precisions and weighted precisions for the category “Animals” as the number of retrieved images increase.

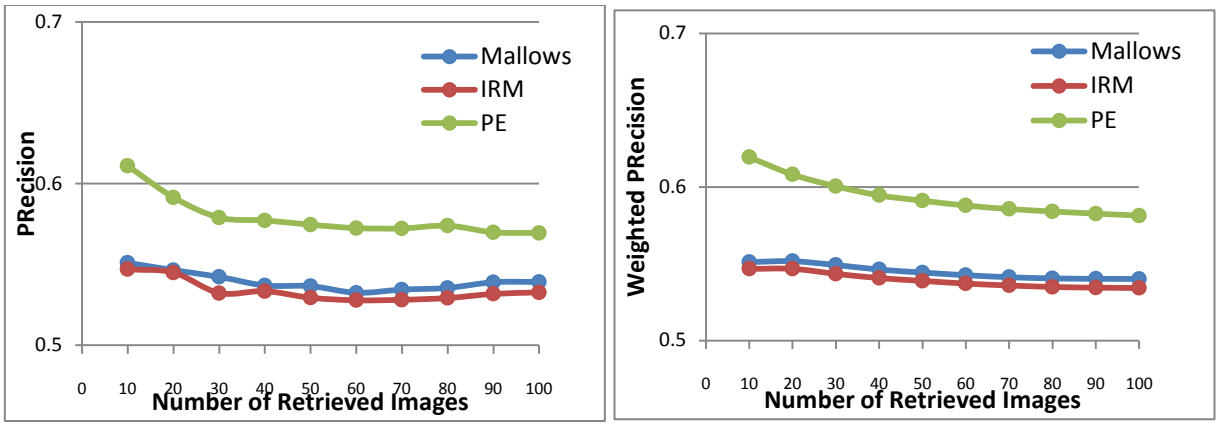


**Figure 18. Average and Weighted Precisions of Category "Animals" with Different Number of Retrieved Images**

From Figure 18, we can see that the average precisions and weighted precisions do not vary much as the number of retrieved images increases. This trend is applicable for other categories of the MIR Flickr dataset as well. Since the number of retrieved images has little effect on the results of this experiment, we analyze only the results of the top 100 retrieved images.

From Figure 17, we can see that overall potential energy distance performs slightly better than does either Mallows distance or IRM. The average precision of potential energy distance is higher than Mallows distance by 0.035 and higher than IRM by 0.037. The weighted precision of potential energy distance is higher than Mallows distance by 0.04 and higher than IRM by 0.042.

We can also see that among all 24 image categories, the average precision of potential energy distance is higher than the other two approaches by over 0.05 for the categories "Bird," "Clouds," "Lake," and "Sea," and lower than the other two approaches for the category "Indoor." For weighted precision, potential energy distance is higher than the other two approaches by over

0.05 for categories "Bird," "Clouds," "Lake," "Sea," and "Sky," and lower than the other two approaches for the category “Indoor.”

In a summary, for both image query and image categorization, potential energy distance performs similarly to Mallows distance and IRM overall, performing better for some categories and worse for other categories.

## 4.4. Speed

In this experiment, image retrieval is performed using an Intel Core i3 2.40 GHz computer and a Linux platform. The process of image segmentation takes most of the experiment time. The sizes of the images in the MIR Flickr dataset are around 500K – 700K. Approximately 45 hours are required to segment 25,000 images and generate image signatures.

The process of image matching is much faster. To compare the image matching performance of potential energy distance with Mallows distance and IRM, we randomly select 1000 images from the MIR Flickr dataset as query images. Image retrieval is performed over all images in the dataset. Table 4 shows the average image matching time of the three approaches.

**Table 4. CPU Time of Different Image Matching Approaches**

| | **Mallows** | **IRM** | **Potential energy** |
|---|---|---|---|
| **Time (Milliseconds)** | **3.46388** | **0.01734** | **0.01618** |

From Table 4, we see that potential energy distance and IRM perform similarly in speed, and both outperform Mallows distance.

## 4.5. Robustness

In this section, extensive experiments are carried out to evaluate the robustness of potential energy distance. We randomly select 6 images as experiment samples, and calculate the rank of the original image and the potential energy distance between the altered image and the original image after various image alternations. Image alternations used in this experiment include different significances of image brightening, image darkening, image blurring with Gaussian filter, image sharpening, image saturation, image random spread, and image pixelization. Figures 19 - 26 list the ranks of the target images and the potential energy distances of different image alternations with different significances. Each color in these figures represents a different experiment sample. Significances are explored from small degrees of alternations to the degrees that potential energy distance approach is not robust any more. In this case, potential energy distance is not robust means a high rank and a large distance from the original image to the altered image.

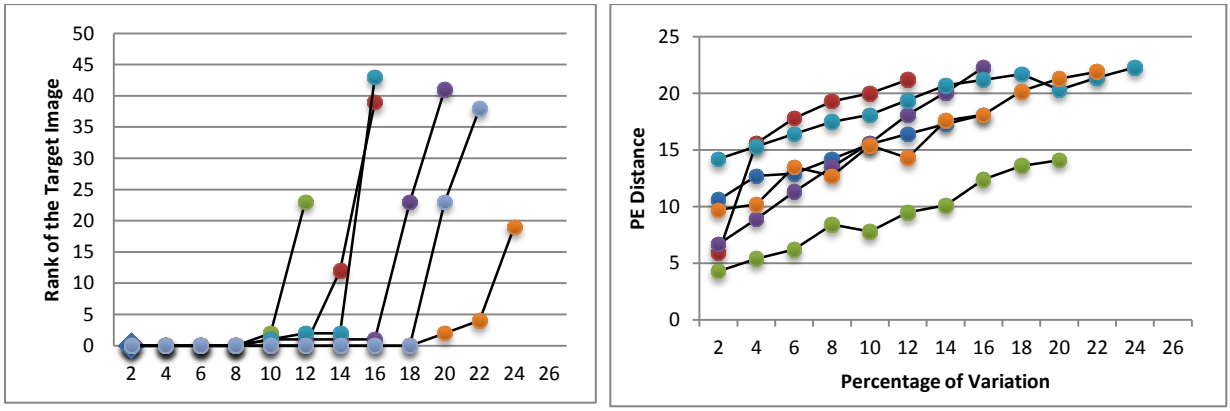


**Figure 19. Rank of the Target Image and PE Distance with Image Brightening**

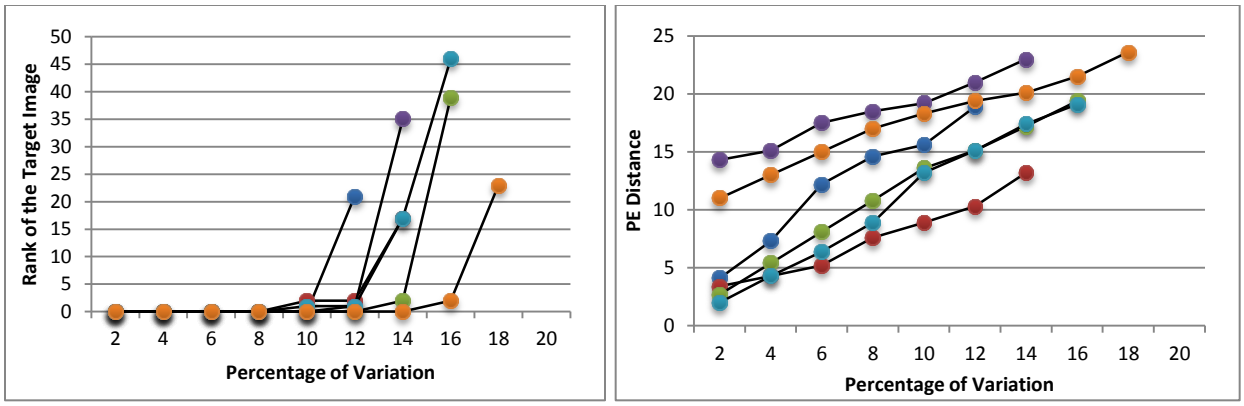


**Figure 20. Rank of the Target Image and PE Distance with Image Darkening**

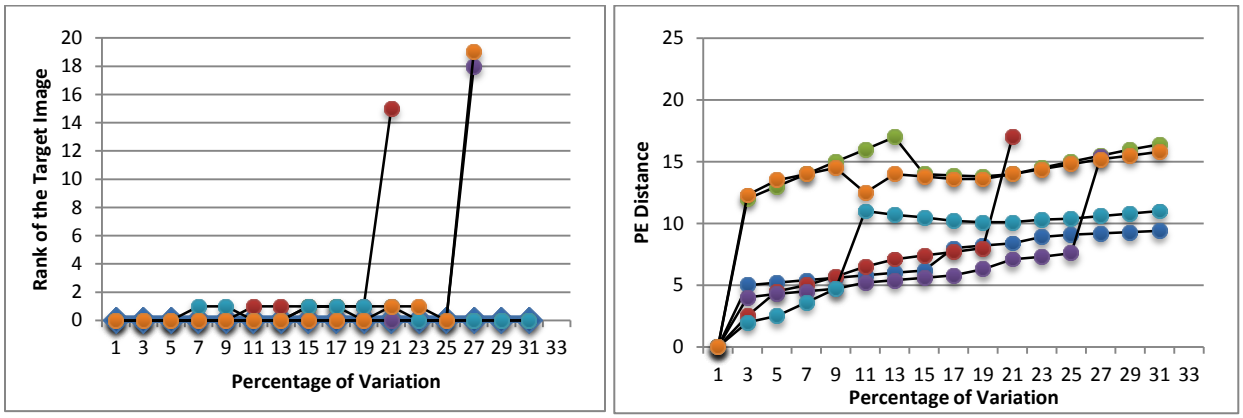


**Figure 21. Rank of the Target Image and PE Distance with Image Blurring**

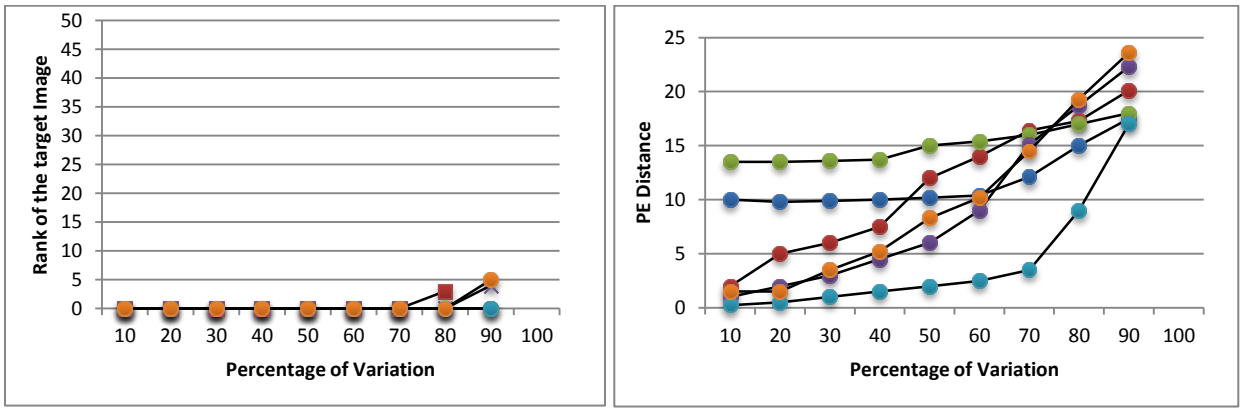


**Figure 22. Rank of the Target Image and PE Distance with Image Sharpening**

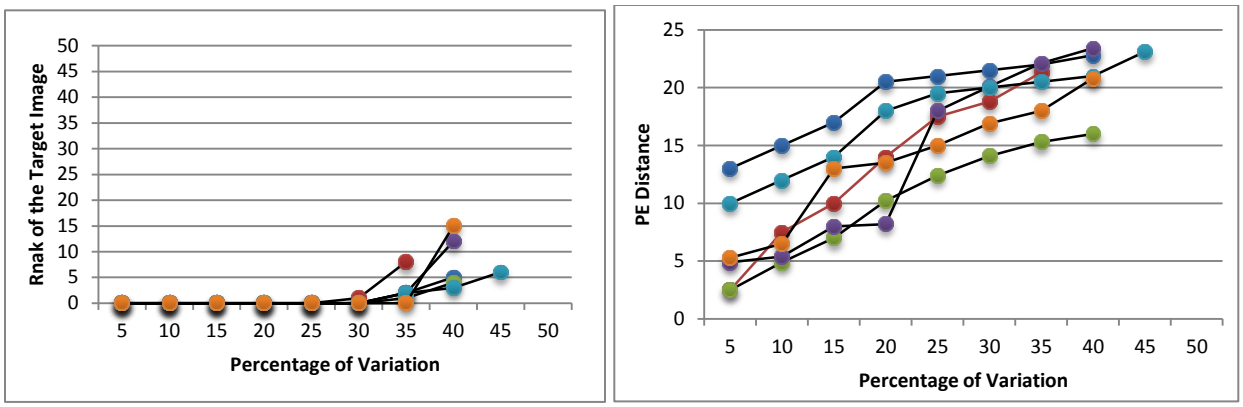


**Figure 23. Rank of the Target Image and PE Distance with More Saturation**

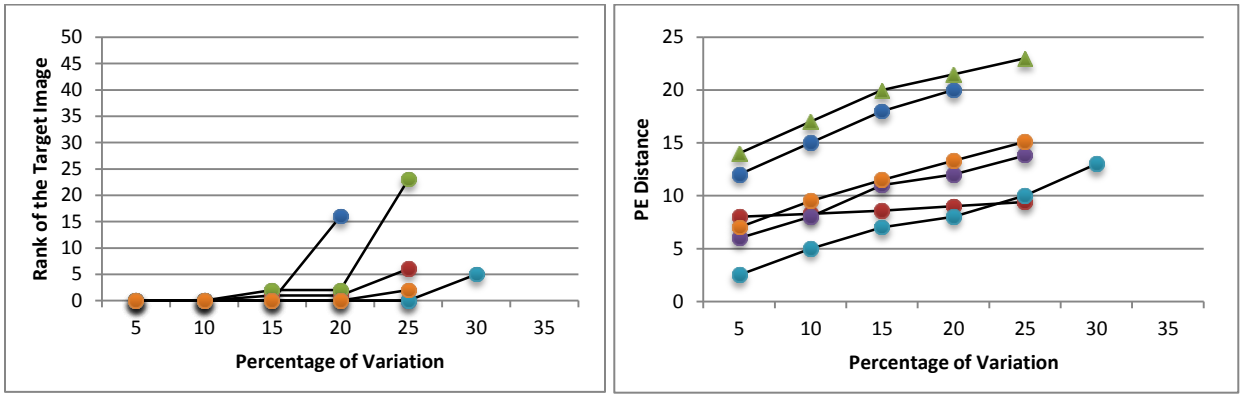


**Figure 24. Rank of the Target Image and PE Distance with Less Saturation**

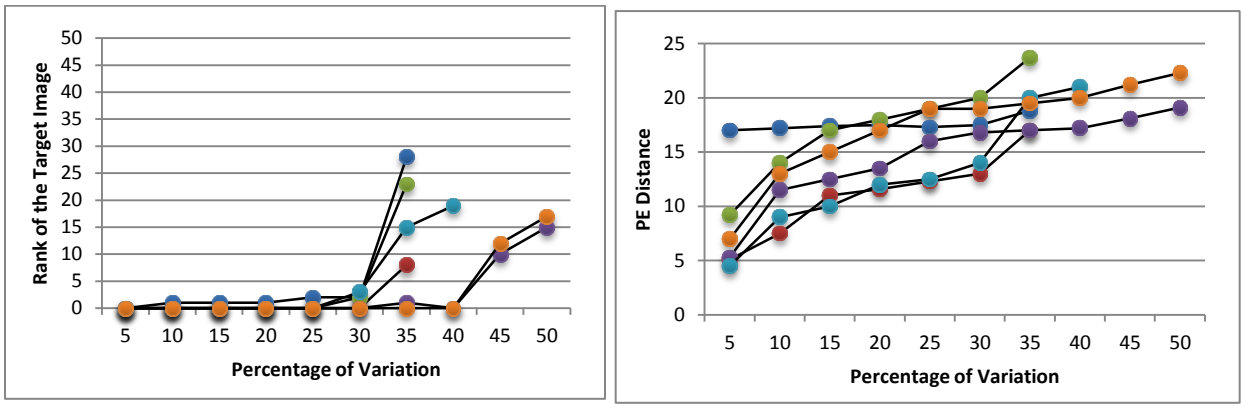


**Figure 25. Rank of the Target Image and PE Distance with Image Random Spread**

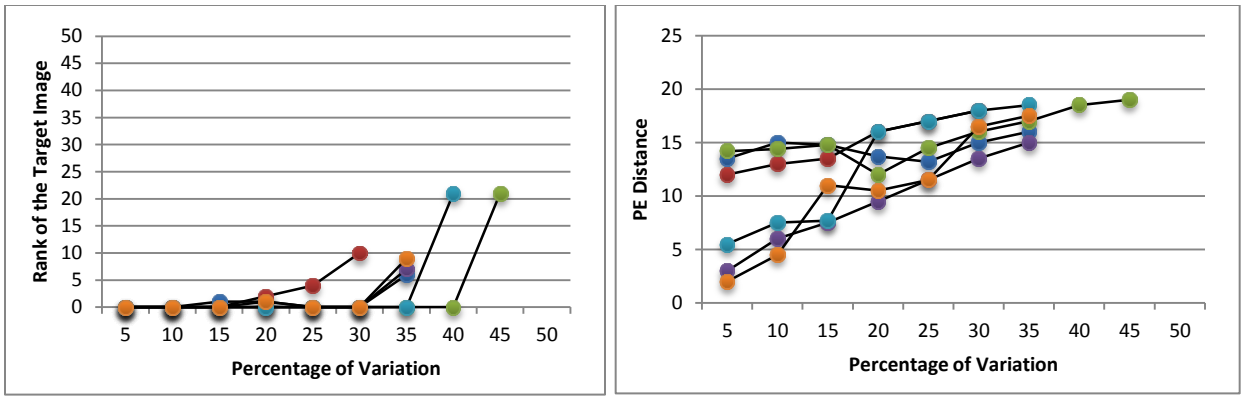


**Figure 26. Rank of the Target Image and PE Distance with Image Pixelization**

From Figures 19 -26, we see that potential energy distance is fairly robust to image alternations. Overall, potential energy distance is robust to approximately 10 percent image brightening, 10 percent darkening, blurring with a 19 * 19 Gaussian filter, 70 percent sharpening, 30 percent more saturation, 15 percent less saturation, random spread by 30 pixels, and pixelization by 30 pixels.

# Chapter 5 Conclusion

As the volume of images on the Internet continues to explode, the variety of visual and semantic content of images grows to cover a lot of different topics. These factors greatly challenge image retrieval. As a result, efficient image retrieval techniques are in great need. Importantly, techniques must keep pace with the incredible growth in images in order to retrieve images with good quality while keeping retrieval speed fast.

In this thesis, we introduce a new similarity measure, namely potential energy distance. Potential energy distance is used to measure the distance between statistical probability distributions. Because potential energy distance assumes that probabilities are independently distributed and does not require optimal matching weights between regions during image matching process, it is fast to compute. Also, potential energy distance is a metric.

To evaluate the performance of potential energy distance, we conducted experiments to evaluate its accuracy, speed, and robustness. Additionally, we compared potential energy distance with two widely used benchmark measures, Mallows distance and IRM.

In our experiment, we used the MIR Flickr dataset that contains 25, 000 images, which can be grouped to 24 categories. During the data pre-processing step, we used segment-based features to construct image signatures. First, we employed MS-A3C approach to segment images into sets of regions and then generated image signatures using the color feature vectors of image regions. The color feature vector of a region was obtained by averaging the LUV colors of all pixels in the region where the weight of a region is the coverage of this region over the whole image

To evaluate the accuracy of potential energy distance, we conducted image query and image categorization experiments. In the image query experiment, we randomly selected 20 images from all 24 categories and implemented image retrieval over all 25,000 images. In the image categorization experiment, we used all 25,000 images as query images and performed image retrieval over all 25,000 images. In these experiments, precision and weighted precision were calculated for each category. Also, because precision and weighted precision may vary depending on the number of retrieved images, we retrieved and analyzed the top 10, top 20, …, top 100 matched images. To evaluate the speed of potential energy distance, we randomly selected 1000 images as query images and implemented image retrieval over all 25, 000 images. In this experiment, the average image matching time of the three different measures was computed. To evaluate the robustness of potential energy distance, we randomly selected a set of query images and applied different image alternations to these images. Ranks of target images and potential energy distances between altered images and target images were then computed.

At last, we answer the research questions proposed in Section 1.4:

For the research question "Q1. Does potential energy distance help improve the accuracy of image retrieval?" The answer is yes for some categories. Our experiment showed that potential energy distance performs better than Mallows distance and IRM for some categories, such as "Bird" and "Clouds," but performs worse than Mallows distance and IRM for other categories, such as “Indoor.”

For the research question "Q2. Does potential energy distance help improve the speed of image retrieval?" The answer is yes. Our experiment showed that potential energy distance is fast to

compute. Moreover, potential energy distance performs similarly to IRM and much faster than Mallows distance.

For the research question "Q3. Is potential energy distance robust for image retrieval?" The answer is yes. Our experiment showed that potential energy distance is fairly robust to different image alternations.

Our work is limited in that we can only conclude that potential energy distance performs closely to IRM and performs much better than Mallows distance in speed, but we cannot determine whether potential energy distance performs better than Mallows distance or IRM in accuracy. The number of images and the number of categories in the dataset we use are too small to compare the retrieval results. Further research that uses larger datasets is needed to evaluate the performance of potential energy distance in image retrieval, such as the Corel dataset, Caltech101 and Caltech256 [70], and MIR Flickr 1 million dataset [69].

# Bibliography


1. Datta, R., et al., *Image retrieval: Ideas, influences, and trends of the new age.* ACM Computing Surveys (CSUR), 2008. **40**(2): p. 5.

2. Smeulders, A.W.M., et al., *Content-based image retrieval at the end of the early years.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2000. **22**(12): p. 1349-1380.

3. Swain, M.J. and D.H. Ballard, *Color indexing.* International journal of computer vision, 1991. **7**(1): p. 11-32.

4. Del Bimbo, A. and P. Pala, *Visual image retrieval by elastic matching of user sketches.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1997. **19**(2): p. 121-132.

5. Wilson, R.C. and E.R. Hancock, *Structural matching by discrete relaxation.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1997. **19**(6): p. 634-648.

6. Wolfson, H.J. and I. Rigoutsos, *Geometric hashing: An overview.* Computational Science & Engineering, IEEE, 1997. **4**(4): p. 10-21.

7. Fagin, R. *Combining fuzzy information from multiple systems*. in *Proceedings of the fifteenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems*. 1996. ACM.

8. Wu, Y., Q. Tian, and T.S. Huang. *Discriminant-EM algorithm with application to image retrieval*. in *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*. 2000. IEEE.

9. Weber, M., M. Welling, and P. Perona, *Unsupervised learning of models for recognition.* Computer Vision-ECCV 2000, 2000: p. 18-32.

10. He, J., et al. *Manifold-ranking based image retrieval*. in *Proceedings of the 12th annual ACM international conference on Multimedia*. 2004. ACM.

11. Vasconcelos, N. and A. Lippman, *A multiresolution manifold distance for invariant image similarity.* Multimedia, IEEE Transactions on, 2005. **7**(1): p. 127-142.

12. He, J., et al. *Mean version space: a new active learning method for content-based image retrieval*. in *Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval*. 2004. ACM.

13. He, X. *Incremental semi-supervised subspace learning for image retrieval*. in *Proceedings of the 12th annual ACM international conference on Multimedia*. 2004. ACM.

14. He, X., W.-Y. Ma, and H.-J. Zhang. *Learning an image manifold for retrieval*. in *Proceedings of the 12th annual ACM international conference on Multimedia*. 2004. ACM.

15. Zhou, D., et al., *Ranking on data manifolds.* Advances in neural information processing systems, 2004. **16**: p. 169-176.

16. Silva, V.d. and J.B. Tenenbaum, *Global versus local methods in nonlinear dimensionality reduction.* Advances in neural information processing systems, 2002. **15**: p. 705-712.

17. Hastie, T., R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction.* 2001.

18. Do, M.N. and M. Vetterli, *Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance.* Image Processing, IEEE Transactions on, 2002. **11**(2): p. 146-158.

19. Iqbal, Q. and J. Aggarwal, *Retrieval by classification of images containing large manmade objects using perceptual grouping.* Pattern recognition, 2002. **35**(7): p. 1463-1479.

20. Zhu, L., et al. *Keyblock: An approach for content-based image retrieval*. in *Proceedings of the eighth ACM international conference on Multimedia*. 2000. ACM.

21. Mathiassen, J., A. Skavhaug, and K. Bø, *Texture similarity measure using Kullback-Leibler divergence between gamma distributions.* Computer Vision—ECCV 2002, 2002: p. 19-49.

22. Pi, M., M.K. Mandal, and A. Basu, *Image retrieval based on histogram of fractal parameters.* Multimedia, IEEE Transactions on, 2005. **7**(4): p. 597-605.

23. Laaksonen, J., M. Koskela, and E. Oja, *PicSOM-self-organizing image retrieval with MPEG-7 content descriptors.* Neural Networks, IEEE Transactions on, 2002. **13**(4): p. 841-853.

24. Wu, G., E.Y. Chang, and N. Panda. *Formulating context-dependent similarity functions*. in *Proceedings of the 13th annual ACM international conference on Multimedia*. 2005. ACM.

25. Natsev, A. and J.R. Smith. *A study of image retrieval by anchoring*. in *Multimedia and Expo, 2002. ICME'02. Proceedings. 2002 IEEE International Conference on*. 2002. IEEE.

26. Theoharatos, C., et al., *A generic scheme for color image retrieval based on the multivariate Wald-Wolfowitz test.* Knowledge and Data Engineering, IEEE Transactions on, 2005. **17**(6): p. 808-819.

27. Edu, Q.Z.Q.W., S.A. Goldman, and W. Yu, *Content-Based Image Retrieval Using Multiple-Instance Learning.*

28. Wang, G., D. Hoiem, and D. Forsyth, *Learning image similarity from flickr groups using fast kernel machines.* 2012.

29. Chechik, G., et al. *An online algorithm for large scale image similarity learning*. in *Proc. NIPS*. 2009. Citeseer.

30. Jiang, S., X. Song, and Q. Huang, *Relative image similarity learning with contextual information for Internet cross-media retrieval.* Multimedia Systems, 2013: p. 1-13.

31. Jin, R. and A.G. Hauptmann. *Using a probabilistic source model for comparing images*. in *Image Processing. 2002. Proceedings. 2002 International Conference on*. 2002. IEEE.

32. Vasconcelos, N. and A. Lippman. *A probabilistic architecture for content-based image retrieval*. in *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*. 2000. IEEE.

33. Vasconcelos, N., *On the efficient evaluation of probabilistic similarity functions for image retrieval.* Information Theory, IEEE Transactions on, 2004. **50**(7): p. 1482-1496.

34. Ning, J., et al., *Interactive image segmentation by maximal similarity based region merging.* Pattern Recognition, 2010. **43**(2): p. 445-456.

35. Androutsos, D., K. Plataniotiss, and A.N. Venetsanopoulos. *Distance measures for color image retrieval*. in *Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on*. 1998. IEEE.

36. Chen, Z. and B. Zhu, *Some Formal Analysis of Rocchio's Similarity-Based Relevance Feedback Algorithm.* Algorithms and Computation, 2000: p. 129-150.

37. Beecks, C., M.S. Uysal, and T. Seidl. *A comparative study of similarity measures for content-based multimedia retrieval*. in *Multimedia and Expo (ICME), 2010 IEEE International Conference on*. 2010. IEEE.

38. Faloutsos, C., et al., *Efficient and effective querying by image content.* Journal of intelligent information systems, 1994. **3**(3): p. 231-262.

39. Ardizzoni, S., I. Bartolini, and M. Patella. *Windsurf: Region-based image retrieval using wavelets*. in *Database and Expert Systems Applications, 1999. Proceedings. Tenth International Workshop on*. 1999. IEEE.

40. Mallows, C., *A note on asymptotic joint normality.* The Annals of Mathematical Statistics, 1972. **43**(2): p. 508-515.

41. Rubner, Y., C. Tomasi, and L.J. Guibas, *The earth mover's distance as a metric for image retrieval.* International Journal of Computer Vision, 2000. **40**(2): p. 99-121.

42. Rubner, Y., C. Tomasi, and L.J. Guibas. *A metric for distributions with applications to image databases*. in *Computer Vision, 1998. Sixth International Conference on*. 1998. IEEE.

43. Wang, J.Z., J. Li, and G. Wiederhold, *SIMPLIcity: Semantics-sensitive integrated matching for picture libraries.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2001. **23**(9): p. 947-963.

44. Ko, B. and H. Byun. *Integrated region-based image retrieval using region's spatial relationships*. in *Pattern Recognition, 2002. Proceedings. 16th International Conference on*. 2002. IEEE.

45. Zhang, R. and Z. Zhang. *Hidden semantic concept discovery in region based image retrieval*. in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*. 2004. IEEE.

46. Jing, F., et al., *An efficient and effective region-based image retrieval framework.* Image Processing, IEEE Transactions on, 2004. **13**(5): p. 699-709.

47. Du, Y. and J.Z. Wang. *A scalable integrated region-based image retrieval system*. in *Image Processing, 2001. Proceedings. 2001 International Conference on*. 2001. IEEE.

48. Chen, Y. and J.Z. Wang, *A region-based fuzzy feature matching approach to content-based image retrieval.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2002. **24**(9): p. 1252-1267.

49. Amores, J., et al. *Boosting contextual information in content-based image retrieval*. in *Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval*. 2004. ACM.

50. Hoiem, D., et al. *Object-based image retrieval using the statistical structure of images*. in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*. 2004. IEEE.

51. Dagli, C. and T.S. Huang. *A framework for grid-based image retrieval*. in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*. 2004. IEEE.

52. Sampat, M.P., et al., *Complex wavelet structural similarity: A new image similarity index.* Image Processing, IEEE Transactions on, 2009. **18**(11): p. 2385-2401.

53. Beecks, C., M.S. Uysal, and T. Seidl. *Signature quadratic form distances for content-based similarity*. in *Proceedings of the 17th ACM international conference on Multimedia*. 2009. ACM.

54. Beecks, C., M.S. Uysal, and T. Seidl. *Signature quadratic form distance*. in *Proceedings of the ACM International Conference on Image and Video Retrieval*. 2010. ACM.

55. Beecks, C., M.S. Uysal, and T. Seidl. *Efficient k-nearest neighbor queries with the signature quadratic form distance*. in *Data Engineering Workshops (ICDEW), 2010 IEEE 26th International Conference on*. 2010. IEEE.

56. Vázquez, P.-P. and J. Marco, *Using Normalized Compression Distance for image similarity measurement: an experimental study.* The Visual Computer, 2011: p. 1-22.

57. Carson, C., et al., *Blobworld: Image segmentation using expectation-maximization and its application to image querying.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2002. **24**(8): p. 1026-1038.

58. Joshi, D., M. Naphade, and A. Natsev. *A greedy performance driven algorithm for decision fusion learning*. in *Image Processing, 2007. ICIP 2007. IEEE International Conference on*. 2007. IEEE.

59. Szekely, G.J., *Potential and kinetic energy in statistics.* Lecture Notes, Budapest Institute, 1989.

60. Bozkaya, T. and M. Ozsoyoglu. *Distance-based indexing for high-dimensional metric spaces*. in *ACM SIGMOD Record*. 1997. ACM.

61. Berretti, S., A. Del Bimbo, and E. Vicario. *Modeling spatial relationships between color sets*. in *Content-based Access of Image and Video Libraries, 2000. Proceedings. IEEE Workshop on*. 2000. IEEE.

62. Assfalg, J., A. Del Bimbo, and P. Pala. *Using multiple examples for content-based image retrieval*. in *Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on*. 2000. IEEE.

63. Li, J., *Agglomerative connectivity constrained clustering for image segmentation.* Statistical Analysis and Data Mining, 2011. **4**(1): p. 84-99.

64. Li, J. and J.Z. Wang, *Studying digital imagery of ancient paintings by mixtures of stochastic models.* Image Processing, IEEE Transactions on, 2004. **13**(3): p. 340-353.

65. Li, J. and J.Z. Wang, *Real-time computerized annotation of pictures.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2008. **30**(6): p. 985-1002.

66. Shirdhonkar, S. and D.W. Jacobs. *Approximate earth mover's distance in linear time*. in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. 2008. IEEE.

67. Carson, C., et al. *Blobworld: A system for region-based image indexing and retrieval*. in *Visual Information and Information Systems*. 1999. Springer.

68. Ma, W.-Y. and B. Manjunath. *Netra: A toolbox for navigating large image databases*. in *Image Processing, 1997. Proceedings., International Conference on*. 1997. IEEE.

69. Huiskes, M.J., B. Thomee, and M.S. Lew. *New trends and ideas in visual concept detection: the MIR flickr retrieval evaluation initiative*. in *Proceedings of the international conference on Multimedia information retrieval*. 2010. ACM.

70. Kohavi, R. and F. Provost, *Glossary of terms.* Machine Learning, 1998. **30**(2-3): p. 271-274.