Perception Based Image Classification

Christopher Henry
James F. Peters, Supervisor
{chenry,jfpeters}@ee.umanitoba.ca
University of Manitoba
Computational Intelligence Laboratory
ENGR E1-526 Engineering & Information Technology Complex
75A Chancellor's Circle
Winnipeg, Manitoba R3T 5V6
Tel.: 204.474.9603
Fax.: 204.261.4639

UM CI Laboratory Technical Report
Number TR-2009-016
May 30, 2009

University of Manitoba

Computational Intelligence Laboratory
URL: http://wren.ee.umanitoba.ca

Project no. 2009-016

Perception Based Image Classification

Christopher Henry
James F. Peters, Supervisor
{chenry,jfpeters}@ee.umanitoba.ca
University of Manitoba
Computational Intelligence Laboratory
ENGR E1-526 Engineering & Information Technology Complex
75A Chancellor's Circle
Winnipeg, Manitoba R3T 5V6
Tel.: 204.474.9603
Fax.: 204.261.4639

CI Laboratory TR-2009-016

May 30, 2009

Abstract

Pattern classification methodologies are present in many systems that we depend on daily. In these systems, classes are created based on human perception of the objects being classified. Thus, it is important to have systems that accurately model human perception. Near set theory provides a framework for measuring the similarity of objects based on features that describe them in much the same way that humans perceive objects. In this paper, we show that the near set approach can be used to classify images. Further, the results presented here suggest that the near set approach can be used in any image classification system. The contribution of this article is a perception based classification of images using near sets.

1 Introduction

The problem addressed in this article is one of reconciling human perception with that of image processing and pattern recognition systems. The term *perception* appears in the literature in many different places with respect to the processing of images. For instance, the term is often used for demonstrating that the performance of methods are similar to results obtained by human subjects (as in [1]), or it is used when the system is trained from data generated by human subjects (as in [2]). Thus, in these examples, a system is considered perceptual if it mimics human behaviour. Another illustration of the use of perception is in the area of semantics with respect to queries [3, 4]. For instance, [4] focuses on queries for 3-D environments, *i.e.*, performing searches of an online virtual environment. Here the question of perception is one of semantics and conceptualization with regard to language and queries. For example, a user might want to search for the tall tree they remembered seeing on one of their visits to a virtual city.

This research work has been funded by Manitoba Hydro grants T137, T247, T260, T270, T277, and by the Natural Sciences & Engineering Research Council of Canada (NSERC) grant 185986, NSERC Postgraduate Doctoral Fellowship PGS-D3, University of Manitoba Faculty of Engineering grant, and Canadian Arthritis Network grant SRI-BIO-05.

Other interpretations of *perception* are tightly coupled to psychophysics, *i.e.* perception based on the relationship between stimuli and sensation. For example, [5] introduces a texture perception model. The texture perception model uses the antagonistic view of the Human Visual System (HVS) in which our brain processes differences in signals received from rods and cones rather than sense signals, directly. An image-feature model of perception has been suggested by Mojsilovic *et al.* [6], where it is suggested that humans view/recall an image by its dominant colours only, and areas containing small, non-dominant colours are averaged by the HVS. Other examples of the term perception defined in the context of psychophysics have also been given [7–13].

Perception as explained by psychologists [14, 15] is similar to the understanding of perception in psychophysics. In a psychologist's view of perception, the focus is more on the mental processes involved rather than interpreting external stimuli. For example, [15] presents an algorithm for detecting the differences between two images based on the representation of the image in the human mind (*e.g.*, colours, shapes, and sizes of regions and objects) rather than on interpreting the stimuli produced when looking at an image. In other words, the stimuli from two images have been perceived and the mind must now determine the degree of similarity.

The view of perception presented in this article combines the basic understanding of perception in psychophysics with a view of perception found in Merleau-Ponty's work [16]. That is, perception of an object (*i.e.*, in effect, our knowledge about an object) depends on information gathered by our senses. The proposed approach to perception is feature-based and is similar to the one discussed in the introduction of [17]. In this view, our senses are likened to probe functions (*i.e.*, mappings of sensations to values assimilated by the mind). A human sense modelled as a probe measures the physical characteristics of objects in our environment. The sensed phyical characteristics of an object are identified with object features. It is our mind that identifies relationships between object feature values to form perceptions of sensed objects [16]. In this article, we show that perception, *i.e.* human perception, can be quantified through the use of near sets by providing a framework for comparing objects based on object descriptions. Objects that have the same appearance (*i.e.*, objects with matching descriptions) are considered *perceptually near each other*. Sets are considered near each other when they have "things" (perceived objects) in common. Specifically, near sets facilitate measurement of similarity between objects based on feature values (obtained by probe functions) that describe the objects. This approach is similar to the way human perceive objects (see, *e.g*, [18]) and as such facilitates pattern classification systems. Much work has been reported in the area of near sets [19–21], which are an outgrowth of the rough set approach to obtaining approximate knowledge of objects that are known imprecisely [22–26].

Pattern classification methodologies can be found in many systems, ranging from photo radar to assembly line manufacturing. In each case, feature vectors are generated from each unknown object being classified. Many of these systems use a supervised learning approach where a training set is employed to learn a decision function for classifying unknown patterns [27]. This article introduces a Nearness Measure (NM) based on near set theory that measures the similarity of pairs of images. Furthermore, the NM is used in a supervised learning environment to classify an unknown image and the results are compared with results obtained using Support Vector Machines (SVMs) [27–29]. The contribution of this article is a perception based classification of images using near sets.

This article is organized as follows: Section 2 gives a brief introduction to near sets with an emphasis on indiscernibility and tolerance relations. Section 3 outlines the steps for combining near set theory with image processing for use in pattern classification. Section 4 provides an overview of SVM and Section 5 presents a comparison of results using using near sets and SVMs for image classification. The work presented in this article is a continuation of recent applications of near set theory reported in [30–34], and the contribution of this work is a step toward perception-based pattern classification.

2 Near sets

Near set theory focuses on sets of perceptual objects with matching descriptions. Specifically, let O represent the set of all objects. The description of an object $x \in O$ is given by

$$\phi(x) = (\phi_1(x), \phi_2(x), \ldots, \phi_i(x), \ldots, \phi_l(x)),$$

where l is the length of the description and each $\phi_i(x)$ is a probe function that describes the object x. Furthermore, we can define a set $\mathbb{F}$ that represents all the probe functions used to describe an object x. Next, a perceptual information system S can be defined as $S = \langle O, \mathbb{F}, \{Val_{\phi_i}\}_{\phi_i \in \mathbb{F}} \rangle$, where $\mathbb{F}$ is the set of all possible probe functions that take as the domain objects in O, and $\{Val_{\phi_i}\}_{\phi_i \in \mathbb{F}}$ is the value range of a function $\phi_i \in \mathbb{F}$. For simplicity, a perceptual system is abbreviated as $\langle O, \mathbb{F} \rangle$ when the range of the probe functions is understood. It is the notion of a perceptual system that is at the heart of the following definitions.

Definition 1 Indiscernibility Relation *Let $\langle O, \mathbb{F} \rangle$ be a perceptual system. For every $\mathcal{B} \subseteq \mathbb{F}$ the indiscernibility relation $\sim_{\mathcal{B}}$ is defined as follows:*

$$\sim_{\mathcal{B}} = \{(x, y) \in O \times O : \parallel \phi(x) - \phi(y) \parallel = 0\},$$

where $\|\cdot\|$ represents the l^2 norm. If $\mathcal{B} = \{\phi\}$ for some $\phi \in \mathbb{F}$, instead of $\sim_{\{\phi\}}$ we write $\sim_{\phi}$.

Defn. 1 is a refinement of the original indiscernibility relation given by Pawlak in 1981 [22]. Using the indiscernibility relation, objects with matching descriptions can be grouped together forming granules of highest object resolution determined by the probe functions in $\mathcal{B}$. This gives rise to an elementary set

$$x_{/\sim_{\mathcal{B}}} = \{x' \in X \mid x' \sim_{\mathcal{B}} x\},$$

defined as a set where all objects have the same description. Similarly, a quotient set is the set of all elementary sets defined as

$$O_{/\sim_{\mathcal{B}}} = \{x_{/\sim_{\mathcal{B}}} \mid x \in O\}.$$

Defn. 1 provides the framework for comparisons of sets of objects by introducing a concept of nearness within a perceptual system. Sets can be considered near each other when they have "things" in common. In the context of near sets, the "things" can be quantified by granules of a perceptual system, *i.e.*, the elementary sets. For practical reasons, the absolute character of $\sim_{\mathcal{B}}$ leads to a weakened relation between sets X, Y where one can find at least one pair of objects $x \in X, y \in Y$ that have matching descriptions. Then we say that X, Y are weakly near each other.

Definition 2 Weak Nearness Relation [35] *Let $\langle O, \mathbb{F} \rangle$ be a perceptual system and let $X, Y \subseteq O$. A set X is weakly near to a set Y within the perceptual system $\langle O, \mathbb{F} \rangle$ $(X \bowtie_{\mathbb{F}} Y)$ iff there are $x \in X$ and $y \in Y$ and there is $\mathcal{B} \subseteq \mathbb{F}$ such that $x \sim_{\mathcal{B}} y$. In the case where sets X, Y are defined within the context of a perceptual system as in Defn 2, then X, Y are weakly near each other*[1].

Let the sets X and Y be near each other in $\langle O, \mathbb{F} \rangle$, *i.e.*, there exists $x \in X, y \in Y, \mathcal{B} \subseteq \mathbb{F}$ such that $x \sim_{\mathcal{B}} y$. Then, as reported in [32], a NM between X and Y is given in (1).

$$NM_{\sim_{\mathcal{B}}} = \frac{\sum_{x_{/\sim_{\mathcal{B}}} \in X_{/\sim_{\mathcal{B}}}} \sum_{y_{/\sim_{\mathcal{B}}} \in Y_{/\sim_{\mathcal{B}}}} \eta\left(x_{/\sim_{\mathcal{B}}}, y_{/\sim_{\mathcal{B}}}\right)}{\max(|X_{/\sim_{\mathcal{B}}}|, |Y_{/\sim_{\mathcal{B}}}|)}, \tag{1}$$

[1] A comparison on the difference between a nearness relation and a weak nearness relation is outside the scope of this paper. For further discussion see [35].

where

$$\eta\left(x_{/\sim_{\mathcal{B}}}, y_{/\sim_{\mathcal{B}}}\right) = \begin{cases} \min(|x_{/\sim_{\mathcal{B}}}|, |y_{/\sim_{\mathcal{B}}}|) & \text{, if } \| \phi(x) - \phi(y) \| = 0, \\ 0 & \text{, otherwise.} \end{cases}$$

As an example of the degree of nearness between two sets, consider Fig. 1 in which each image consists of two sets of objects, X and Y, that are subsets of the universe of objects O. Each colour in the figures corresponds to an elementary set where all the objects in the class share the same description. The idea behind Eq. 1 is that the nearness of sets in a perceptual system is based on the cardinality of equivalence classes that they share. Thus, the sets in Fig. 1(a) are closer (more near) to each other in terms of their descriptions than the sets in Fig. 1(b).

(a) (b)

Figure 1: Example of degree of nearness between two sets: (a) High degree of nearness, and (b) low degree of nearness.

2.1 Tolerance relation

When dealing with perceptual objects (especially, components in images), it is sometimes necessary to relax the equivalence condition of Defn. 1 to facilitate observation of associations in a perceptual system. This variation is called a tolerance relation and is given in Defn. 3.

Definition 3 Tolerance Relation *Let $\langle O, \mathbb{F} \rangle$ be a perceptual system and let $\epsilon \in \mathbb{R}$ (set of all real numbers). For every $\mathcal{B} \subseteq \mathbb{F}$ the tolerance relation $\cong_{\mathcal{B}}$ is defined as follows:*

$$\cong_{\mathcal{B},\epsilon} = \{(x, y) \in O \times O : \| \phi(x) - \phi(y) \| \leq \epsilon\}.$$

If $\mathcal{B} = \{\phi\}$ for some $\phi \in \mathbb{F}$, instead of $\cong_{\{\phi\}}$ we write $\cong_{\phi}$. Further, for notational convince, we will write $\cong_{\mathcal{B}}$ instead of $\cong_{\mathcal{B},\epsilon}$ with the understanding that ϵ is inherent to the definition of the tolerance relation.

As in the case with the indiscernibility relation, a tolerance class can be defined as

$$x_{/\cong_{\mathcal{B}}} = \{y \in X \mid y \cong_{\mathcal{B}} x\}. \tag{2}$$

Note, Defn. 3 does not uniquely partition O (*i.e.* an object can belong to more than one class) which is why Eq. 2 is called a tolerance class instead of an elementary set. In addition, each pair of objects x, y in a tolerance class $x_{/\cong_{\mathcal{B}}}$ must satisfy the condition $\| \phi(x) - \phi(y) \| \leq \epsilon$. Next, a quotient set for a given a tolerance relation is the set of all tolerance classes and is defined as

$$O_{/\cong_{\mathcal{B}}} = \{x_{/\cong_{\mathcal{B}}} \mid x \in O\}.$$

As was the case with the equivalence relation, tolerance classes reveal relationships in perceptual systems leading to the definition of a tolerance nearness relation.

Definition 4 Weak Tolerance Nearness Relation *[36]*
Let $\langle O, \mathbb{F}\rangle$ be a perceptual system and let $X, Y \subseteq O, \epsilon \in \mathbb{R}$. The set X is perceptually near to the set Y within the perceptual system $\langle O, \mathbb{F}\rangle$ ($X \underline{\bowtie}_{\mathbb{F}} Y$) iff there exists $x \in X$, $y \in Y$ and there is a $\phi \in \mathbb{F}, \epsilon \in \mathbb{R}$ such that $x \cong_{\mathcal{B}} y$. If a perceptual system is understood, then we say shortly that a set X is perceptually near to a set Y in a weak tolerance sense of nearness.

Similar to Eq. 1, a NM under the tolerance relation is given as

$$NM_{\cong_{\mathcal{B}}} = \sum_{x_{/\cong_{\mathcal{B}}} \in X_{/\cong_{\mathcal{B}}}} \sum_{y_{/\cong_{\mathcal{B}}} \in Y_{/\cong_{\mathcal{B}}}} \frac{\xi\left(x_{/\cong_{\mathcal{B}}}, y_{/\cong_{\mathcal{B}}}\right)}{\max(|x_{/\cong_{\mathcal{B}}}|, |y_{/\sim_{\mathcal{B}}}|)}, \tag{3}$$

where

$$\xi\left(x_{/\cong_{\mathcal{B}}}, y_{/\cong_{\mathcal{B}}}\right) = \begin{cases} \min(|x_{/\cong_{\mathcal{B}}}|, |y_{/\cong_{\mathcal{B}}}|) & \text{, if } \parallel \phi(x) - \phi(y) \parallel \leq \epsilon, \\ 0 & \text{, otherwise.} \end{cases}$$

Notice the subtle difference between the two nearness measures. Since objects can belong to more than one tolerance class, the denominator of Eq. 3 has moved inside the summation. Similarly, Eq.'s 1 & 3 are equivalent when $\epsilon = 0$.

The following simple example highlights the need for a tolerance relation as well as demonstrates the construction of tolerance classes from real data. Consider Table 1 that contains 20 objects with $|\phi(x_i)| = 1$. Letting $\epsilon = 0.1$ gives the following tolerance classes:

$$\begin{aligned} X_{/\cong_{\mathcal{B}}} = \; & \{\{x_1, x_8, x_{10}, x_{11}\}, \{x_1, x_9, x_{10}, x_{11}, x_{14}\}, \\ & \{x_2, x_7, x_{18}, x_{19}\}, \\ & \{x_3, x_{12}, x_{17}\}, \\ & \{x_4, x_{13}, x_{20}\}, \{x_4, x_{18}\}, \\ & \{x_5, x_6, x_{15}, x_{16}\}, \{x_5, x_6, x_{15}, x_{20}\}, \\ & \{x_6, x_{13}, x_{20}\}\} \end{aligned}$$

Observe that each object in a tolerance class satisfies the condition $\parallel \phi(x) - \phi(y) \parallel \leq \epsilon$, and that almost all of the objects appear in more than one class. Moreover, there would be twenty classes if the indiscernibility relation was used since there are no two objects with matching descriptions.

3 Near Sets and Image Classification

Near set theory can be used to determine the nearness between two images. The nearness measure can be considered a feature value as defined in pattern classification literature (see, *e.g.*, [27]). The following sections describe an approach for applying near set theory to images.

3.1 Image processing

Briefly, this section defines some image processing notation. Let $M, N \in \mathbb{N}$ respectively denote the quantities width and height, and let the mathematical representation of an image using the grayscale colour model

Table 1: Tolerance Class Example

x_i	$\phi(x)$	x_i	$\phi(x)$	x_i	$\phi(x)$	x_i	$\phi(x)$
x_1	.4518	x_6	.6943	x_{11}	.4002	x_{16}	.6079
x_2	.9166	x_7	.9246	x_{12}	.1910	x_{17}	.1869
x_3	.1398	x_8	.3537	x_{13}	.7476	x_{18}	.8489
x_4	.7972	x_9	.4722	x_{14}	.4990	x_{19}	.9170
x_5	.6281	x_{10}	.4523	x_{15}	.6289	x_{20}	.7143

be defined as $f : \{1, \ldots, M\} \times \{1, \ldots, N\} \longrightarrow [0, 255]$. Similarly, let a subimage f_s of f be defined as $f_s : \{p, \ldots, P\} \times \{q, \ldots, Q\} \longrightarrow [0, 255]$ where $p \leq P \leq M$ and $q \leq Q \leq N$. Furthermore, given an image f, the probability p_i of a pixel taking on a value $i \in [0, 255]$ is given by $p_i = T_i/T$, where T_l is the count of grey level l in f, and $T = M \times N$.

3.2 Information content

Shannon introduced entropy as a measure of the amount of information gained by receiving a message from a finite codebook of messages [37]. The idea was that the gain of information from a single message is proportional to the probability of receiving the message. Thus, receiving a message that is highly unlikely gives more information about the system than a message with a high probability of transmission. Formally, let the probability of receiving a message i of n messages be p_i, then the information gain of a message can be written as

$$\Delta I = \log(1/p_i) = -\log(p_i), \tag{4}$$

and the entropy of the system is the expected value of the gain and is calculated as

$$H = -\sum_{i=1}^{n} pi \log(p_i).$$

However, as reported in [37], Shannon's definition of entropy suffers from three things: it is undefined when $p_i = 0$; that in practise the information gain (whether probable or un-probable) should lie in the interval $[0, 1]$ and not at the limits (which is the case when using Eq. 4); and that a statistically better measure of ignorance is 1 - p_i rather than $1/p_1$. As a result, [37] lists the following desirable properties of an entropic function:

P1: $\Delta I(p_i)$ is defined at all points in $[0, 1]$.

P2: $\lim_{p_i \to 0} \Delta I(p_i) = \Delta I(p_i = 0) = k_1, k_1 > 0$ and finite.

P3: $\lim_{p_i \to 1} \Delta I(p_i) = \Delta I(p_i = 1) = k_2, k_2 > 0$ and finite.

P4: $k_2 < k_1$.

P5: With increase in p_i, $\Delta I(p_i)$ decreases exponentially.

P6: $\Delta I(p)$ and H, the entropy, are continuous for $0 \leq p \leq 1$.

P7: H is maximum when all p_i's are equal, *i.e.* $H(p_1, \ldots, p_n) \leq H(1/n, \ldots, 1/n)$.

Keeping these properties in mind, [37] defines the gain in information from an event as

$$\Delta I(p_i) = e^{(1-p_i)},$$

which gives the entropy as

$$H = \sum_{i=1}^{n} p_i e^{(1-p_i)}.$$

3.3 Example of near images

The nearness of two images can be discovered by partitioning each of the images into subimages and letting these represent objects in a perceptual system, *i.e*, let the sets X and Y represent the two images to be compared where each set consists of the subimages obtained by partitioning the images. Then, the set of all objects in the perceptual system is given by $O = X \cup Y$. Objects in this system can be described by probe functions that operate on images. Simple examples include average colour, or maximum intensity (see, *e.g.*, [38] for other examples of image probe functions). The results presented in this article use the probe functions $\mathcal{B} = \{H(f_s), \text{Avg}(f_s)\}$, where $H(f_s)$ is Pal's entropy of a subimage, and $\text{Avg}(f_s)$ is the average grayscale of a subimage. Average grayscale is used in addition to Pal's entropy to differentiate between areas in an image that have the same information content. For example, a subimage that consists of all black pixels produces the same value of entropy as a subimage that contains all white pixels (or rather any subimage of uniform intensity).

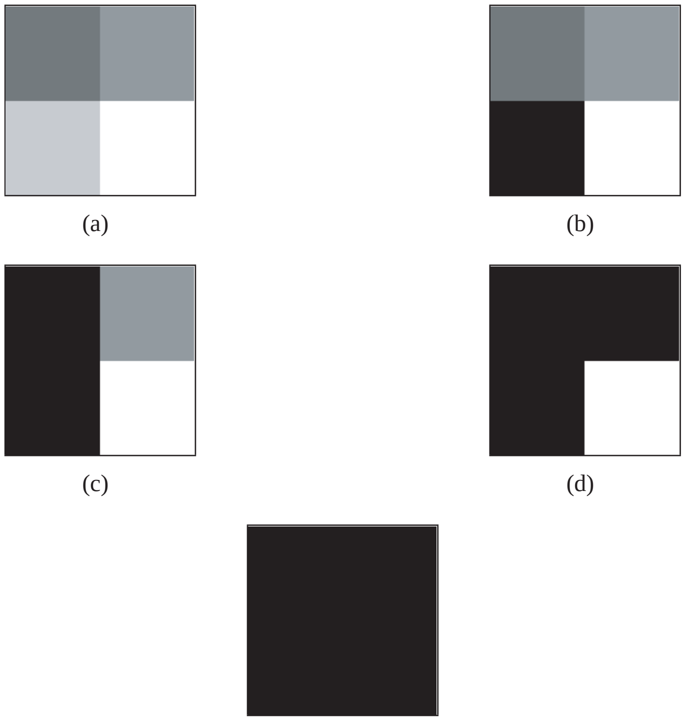
Our first example of near images is given in Fig. 2 where Fig. 2(a) is being compared first to itself and then to Fig.'s 2(b)-2(e). Each image is a Bitmap of size 200×200, each coloured square has dimensions 100×100, and the size of each subimage is 10×10. The NMs were calculated using both the indiscernibility relation (Eq. 1) and the tolerance relation (Eq. 3). Notice that in both cases the NMs are the same due to a small choice of ϵ. In this case, ϵ would have to be much larger than 1 to produce a different NM since the grey levels are not close to each other. Also note, the values range from 1, the case of the image being compared to itself, to 0, the case of the images being completely different.

Our next example provides a visual representation of both equivalence and tolerance classes. Fig. 3 consists of images from the Berkeley Segmentation Dataset [39] and the Leaves Dataset [40], which are also used to obtain the results presented in Sect. 5. Next, Fig. 4 consists of images depicting the equivalence and tolerance classes created from Fig 3(a). Fig. 4(a) was created using Eq. 1 with $\mathcal{B} = \text{Avg}(f_s)$ and a window size of 5 pixels, and each grey level represents a different class. Similarly, Fig. 4(b) shows the number of classes each subimage belongs to, and was created using Eq. 3 with $\epsilon = 0.1$ and a window size of 10 pixels. Notice that it is difficult to display the different classes under the tolerance relation, since each object can belong to more than one class; however, it would look similar to Fig. 4(a) except that each subimage would have multiple colours designating class membership.

Finally, Fig. 5 is a plot of NM values comparing the nearness of Fig.'s 3(a) & 3(b) and Fig.'s 3(a) & 3(c) for $\epsilon = 0, 0.01, 0.05, 0.1$ (note, the indiscernibility relation is used for $\epsilon = 0$). Observe that the two leaf images produce a higher NM than Fig. 3(a) and the Berkeley image because the leaf images produce objects that have more in common in terms of their descriptions (using the probe functions in $\mathcal{B}$). These results match our perception of the similarity between these three images.

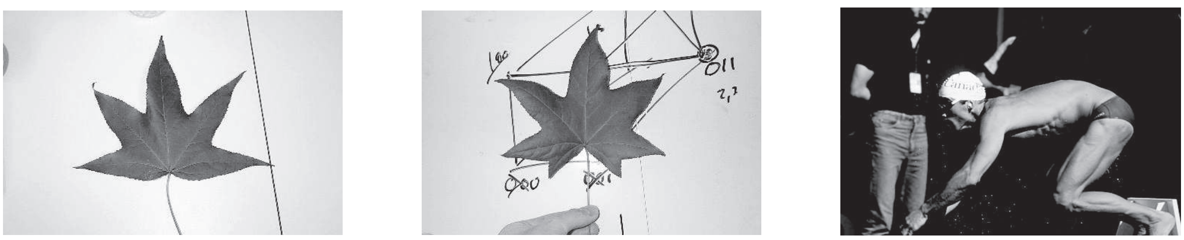
3.4 Algorithms

This section outlines the steps required to calculate the nearness of images using the relations outlined in Sect. 2. Starting with the indiscernibility relation, the first step is to calculate the quotient set (equivalence classes) of each image. This process is given in Alg. 1 and is accomplished by assigning a label, x_c, to each object (subimage) and calculating $\phi(x_c)$. Then the subimages are grouped together based on their



(a) (b) (c) (d) (e)

Figure 2: Example of NM comparing first image to the remaining three: (a) Test pattern for comparison (note, $NM_{\sim_{\mathcal{B}}} = NM_{\cong_{\mathcal{B}}} = 1$ when compared to itself), (b) $NM_{\sim_{\mathcal{B}}} = NM_{\cong_{\mathcal{B}}} = 0.75$, (c) $NM_{\sim_{\mathcal{B}}} = NM_{\cong_{\mathcal{B}}} = 0.5$, (d) $NM_{\sim_{\mathcal{B}}} = NM_{\cong_{\mathcal{B}}} = 0.25$, and (e) $NM_{\sim_{\mathcal{B}}} = NM_{\cong_{\mathcal{B}}} = 0$.

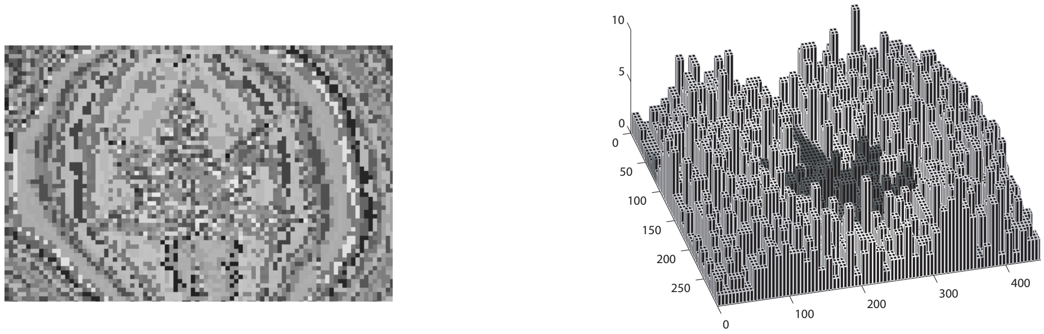


(a) (b) (c)

Figure 3: Samples from image databases: (a), (b) Leaves Dataset [40], and (c) Berkeley Segmentation Dataset [39].

descriptions using the indiscernibility relation. This creates a new partition of the image (called the quotient set) based on object descriptions. Once the quotient set of each image has been determined, the degree of nearness between two sets can be calculated by comparing their elementary sets. This process is described by Alg. 2

The creation of a quotient set under the tolerance relation requires four main tasks. The first step (given in Alg. 3) simply creates a set of objects X from an input image. The next step (given in Alg. 4) is a "first pass" over the set of objects, where the goal is to find for each $x \in X$, a set X' consisting of all the objects in X for which the condition $\| \phi(x) - \phi(x') \| \leq \epsilon$ is satisfied. In other words, X' consists of objects where the only constraint is that all elements must satisfy the tolerance relation with x (and not the rest). The output of



(a) (b)

Figure 4: Examples showing visualization of equivalence and tolerance classes obtained from image Fig 3(a): (a) Equivalence classes created using $\mathcal{B} = \text{Avg}(f_s)$, and (b) plot showing the number of classes a subimage belongs to under the tolerance relation.

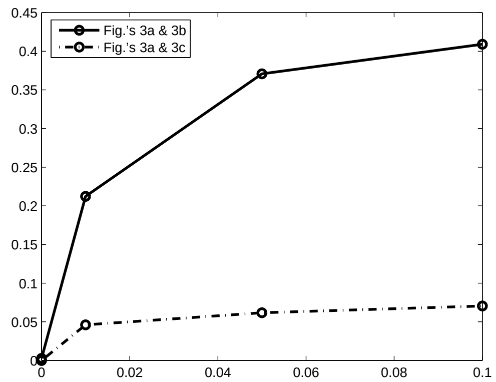


Figure 5: Plot showing NM values comparing Fig.'s 3(a) & 3(b) and Fig.'s 3(a) & 3(c) for $\epsilon = 0, 0.01, 0.05, 0.1$

the algorithm is a collection of all the sets X', *i.e.* $X_c = \bigcup\{X'\}$. The next step involves going through the classes in the set X_c and creating tolerance classes. Thus, for each $\{x_c\} \in X_c$, the goal is to create tolerance classes under which the tolerance relation holds for each unordered pair of elements in the class. This step is shown in Alg. 5. Finally, the last step is to remove any duplicate classes in the quotient set obtained from Alg. 5.

As an example, of Alg.'s 4 & 5, consider again the sample data given in Table 1. Using this data, the

Algorithm 1: Algorithm for calculating equivalence classes

Input : f (image), M (image width), N (image height), γ^2 (area of subimage), $\mathcal{B}$ (set of probe functions for object description)
Output: $X_{/\sim_\mathcal{B}}$ (set of equivalence classes for image f)

```
X_{/~B} ← ∅;
c ← 0;
for (q = 1; q ≤ N; q+ = γ) do
    for (p = 1; p ≤ M; p+ = γ) do
        Q ← min(q + γ − 1, M);
        P ← min(p + γ − 1, N);
        Define subimage f_s over domain {{p, ..., P} × {q, ..., Q}};
        Assign f_s the object label x_c;
        found ← false;
        for (x_{/~B} ∈ X_{/~B}) do
            if x_c ~B x_{/~B} then
                found ← true;
                x_{/~B} ← x_{/~B} ∪ x_c;
            end
        end
        if !found then
            X_{/~B} ← X_{/~B} ∪ {x_c};
        end
        c ← c + 1;
    end
end
```

Algorithm 2: Algorithm for calculating the degree of nearness between two sets

Input : $X_{/\sim_\mathcal{B}}, Y_{/\sim_\mathcal{B}}$ (quotient set of each image)
Output: $NM_{\sim_\mathcal{B}}$

```
NM_{~B} ← 0;
for (x_{/~B} ∈ X_{/~B}) do
    for (y_{/~B} ∈ Y_{/~B}) do
        if x_{/~B} ~B y_{/~B} then
            NM_{~B} ← NM_{~B} + min(|x_{/~B}|, |y_{/~B}|);
        end
    end
end
NM_{~B} ← NM_{~B} / max(|X_{/~B}|, |Y_{/~B}|);
```

output of Alg. 4 is

$$\begin{aligned} X_c \quad = \quad & \{\{x_1, x_8, x_9, x_{10}, x_{11}, x_{14}\}, \\ & \{x_2, x_7, x_{18}, x_{19}\}, \\ & \{x_3, x_{12}, x_{17}\}, \\ & \{x_4, x_{13}, x_{18}, x_{20}\}, \\ & \quad \vdots \\ & \{x_{20}, x_4, x_5, x_6, x_{13}, x_{15}\}\} \end{aligned}$$

Algorithm 3: Tolerance class creation step 1

Input : f (image), M (image width), N (image height), γ^2 (area of subimage)
Output: X (set of objects)

```
X ← ∅;
c ← 0;
for (q = 1; q ≤ N; q+ = γ) do
    for (p = 1; p ≤ M; p+ = γ) do
        Q ← min(q + γ − 1, M);
        P ← min(p + γ − 1, N);
        Define subimage f_s over domain {{p, ..., P} × {q, ..., Q}};
        Assign f_s the object label x_c;
        X ← X ∪ x_c;
        c ← c + 1;
    end
end
```

Algorithm 4: Tolerance class creation step 2

Input : X (set of objects), ϵ
Output: X_c (consisting of sets where the all elements satisfy the tolerance relation with the first element)

```
X_c ← ∅;
for (x ∈ X) do
    x_c ← x;
    for (x' ∈ X) do
        if x ≠ x' and ‖ φ(x) − φ(x') ‖≤ ε then
            x_c ← x_c ∪ x';
        end
    end
    X_c ← X_c ∪ {x_c};
end
```

Similarly, the output of Alg. 5 is

$$
\begin{aligned}
X_{/\cong_{\mathcal{B}}} \quad = \quad & \{\{x_8, x_1, x_{10}, x_{11}\}, \{x_9, x_1, x_{10}, x_{11}, x_{14}\}, \\
& \{x_7, x_2, x_{18}, x_{19}\}, \\
& \{x_{12}, x_3, x_{17}\}, \\
& \{x_{13}, x_4, x_{20}\}, \{x_{18}, x_4\}, \\
& \qquad \vdots \\
& \{x_4, x_{20}, x_{13}\}, \\
& \{x_5, x_{20} x_6, x_{15}\}\}
\end{aligned}
$$

Notice that the order of the elements is the order they are placed in the class by the algorithm. In addition, the output of Alg. 5 will intentionally always produce duplicate classes in order to identify every tolerance class.

Algorithm 5: Tolerance class creation step 3

Input : X_c, ϵ
Output: $X_{/\cong_\mathcal{B}}$ (a quotient set of X containing duplicates classes)
$X_{/\cong_\mathcal{B}} \leftarrow \emptyset$;
for *($\{x_c\} \in X_c$)* **do**
  $x_C \leftarrow x_c$ (used for comparison of objects);
  $x_f \leftarrow$ first element of x_c;
  $x_c \leftarrow x_c \backslash x_f$ (remove the first element from x_c);
  while *($|x_c| > 0$)* **do**
    $x_f \leftarrow$ first element of x_c;
    $x_{c/\cong_\mathcal{B}} \leftarrow x_f$;
    $x_c \leftarrow x_c \backslash x_f$;
    for *($x \in x_C$)* **do**
      Add x to $x_{c/\cong_\mathcal{B}}$ if it is within ϵ of all members of $x_{c/\cong_\mathcal{B}}$;
      Remove x from x_c if it was added to $x_{c/\cong_\mathcal{B}}$;
    end
    $X_{/\cong_\mathcal{B}} \leftarrow X_{/\cong_\mathcal{B}} \cup x_{c/\cong_\mathcal{B}}$;
  end
end

4 Support Vector Machines

Support vector machine (SVMs) map input vectors into a high-dimensional feature space via a non-linear mapping chosen a priori [28]. SVMs are an instance of a popular kernel method for deterministic pattern classification (see, *e.g.*, [27, 29]). In practice, SVMs provide a supervised learning technique requiring training data where the central concept is to find the widest margin in a d-dimensional space between the data belonging to two classes. The data lying on the edge of this margin are called the support vectors and are used to classify the test data.

Formally, the set of training data is given as $T_n = \{\mathbf{x}_i, y_i\}$, $i = 1, \dots, n$, where y_i is the class label and is given by $y_i \in \{-1, 1\}$, and $\mathbf{x}_i \in \mathbb{R}^d$. Assuming that the training data is lineraly seperable, there exists a hyperplane that seperates the data such that the points $\mathbf{x}$ lying on the hyperplane safisfy $\mathbf{w}^T\mathbf{x} + w_0 = 0$, and all $x_i \in T_n$ satisfy

$$y_i(\mathbf{w}^T\mathbf{x}_i + w_0) > 0, \tag{5}$$

where $\mathbf{w}$ is the normal vector of the separating hyperplane. Again using the assumption that the data is linearly seperable, a margin can always be found around the seperating hyperplane representing a "dead zone" in which no training data can be found. As a result, Eq. 5 can be redefined as

$$\begin{aligned} & y_i(\mathbf{w}^T\mathbf{x}_i + w_0) \geq b, \\ \equiv \quad & y_i(\mathbf{w}^T\mathbf{x}_i + w_0) \geq 1. \end{aligned}$$

Next, let us define the support vectors as those points which lie on the edge of the dead zone, *i.e.*, $\{x_i \mid y_i(\mathbf{w}^T\mathbf{x}_i + w_0) = 1\}$. Further, recall that the distance between a point $\mathbf{x}$ and a hyperplane is given as $|\mathbf{w}^T\mathbf{x} + w_0| \;/\; \|\mathbf{w}\|$. Consequently, the distance from the support vectors on either side of the seperating hyper plane is $1 \;/\; \|\mathbf{w}\|$, which gives a margin of $2 \;/\; \|\mathbf{w}\|$. Moreover, the maximum margin can be found by minimizing $0.5 \;\|\mathbf{w}\|^2$ subject to the constraint $y_i(\mathbf{w}^T\mathbf{x}_i + w_0) \geq 1$.

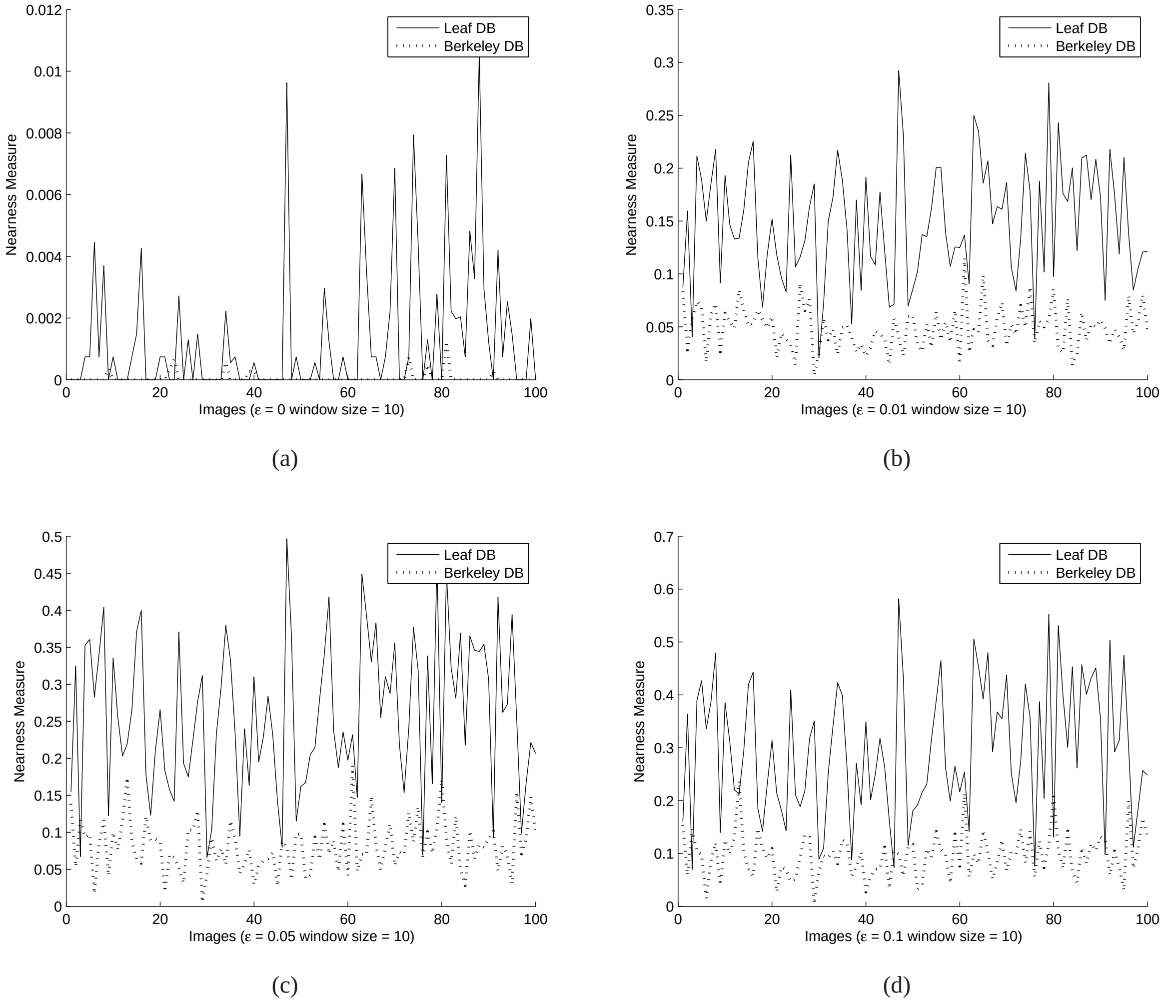


Figure 6: Samples from image databases: (a), (b) Leaves Dataset [40], and (d) Berkeley Segmentation Dataset [39].

This problem can be reformulated in terms of Lagrange multipliers written as

$$L_p = \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_{j=1}^{n} \alpha\{y_i(\mathbf{w}^T\mathbf{w}_j + w_0 - 1\},$$

where now we are minimizing L_p with respect to $\mathbf{w}$, w_0, while also requiring the derivatives of L_p with respect to α_i vanish subject to $\alpha_i \geq 0$ [29]. However, we can solve the dual formulation of this problem and instead maximize

$$L_D = \sum_{j=1}^{n} -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\alpha_j y_i y_y \mathbf{x}_i^T\mathbf{x}_j, \tag{6}$$

subject to $\sum_{j=1}^{n} \alpha_j y_y = 0$, and $\alpha_i \geq 0$ [29]. Similarly, Eq. 6 is also maximized in the case of non-linearly seperable data, except that the constraints are now $\sum_{j=1}^{n} \alpha_j y_y = 0$, and $0 \leq \alpha_i \leq \gamma$, where γ is a penalty assigned to errors. Lastly, it is important to note that Eq. 6 is written in terms of an inner product that allows the definition of a kernel function (this is important for non-linear decision boundaries). For example, a linear kernel could be defined as $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T\mathbf{x}_j$. The two kernels used in this paper are:

- Polynomial: $K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \mathbf{x}_j + 1)^p$, and
- Gaussian: $K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\sigma^2}$.

5 Results

Table 2: Tolerance Class Example

# of training images	# of test images	Alg.	% correctly classified	Alg.	% correctly classified	Alg.	% correctly classified
4	196	NM	93.4	SVM	88.3	SVM	37.8
10	190		90.0	(Poly.)	87.9	(Gaus.)	44.7
14	186		90.9		88.2		57.5
20	180		92.2		87.8		60.6
24	176		92.6		86.9		65.9

This section presents results obtained using the NM to observe similarity in pairs of images. First, we present evidence that demonstrates the NM is up to the task of classifying images. Next, the NM is used to classify images and the results are compared to those obtained from a SVM for a traditional two class classification problem.

5.1 NM for classification

The plot given in Fig. 5 suggests that the NM would be useful in classification of images. To investigate this property further, we have used the NM measure to compare the nearness of images from the Berkeley Segmentation Dataset [39] and from the Leaves Dataset [40] (both freely available online). Specifically, the image in Fig. 3(a) is compared to 200 images (100 from both the leaves and Berkeley datasets, respectively). The results of these comparisons are given in Fig. 6. Note, the number of pixels in the leaf images were decimated by a factor of 4 to be closer in size to the Berkeley images, *i.e.*, their dimension was reduced from 896×592 to 448×296. Further, the measures presented in Eq.'s 1 & 3 were weighted by the size of the indiscernibility/tolerance class. Thus, larger classes contribute more to the NM than smaller ones. Lastly, as was mentioned above, the probe functions selected were $\mathcal{B} = \{H(f_s), \text{Avg}(f_s)\}$, and the size of the subimage in all tests was 10×10.

Notice in each plot, the NM nearness measure associated with comparison between leaf images is (for the most part) larger than that of the NMs associated with comparison between Fig. 3(a) and the Berkeley images. These results match our perception of the images in both datasets, *i.e.*, given only the ability to describe the images using Pal's entropy and average grayscale, we would associate the first leaf image with the rest of the images in the leaf database rather than those in the Berkeley dataset. However, there are exceptions. For example, image 76 (see Fig. 7) produces a very low NM for $\epsilon = 0.01$. In this case, it is clear that (using average grayscale and entropy) these two images are not similar even though they both contain leaves. Likewise, there are also images in the Berkeley dataset that can produce a high NM given $\mathcal{B}$. Lastly, observe that NM values increase with ϵ. Again, this matches our intuition inasmuch as more similarities will be observed between images if the standard for comparison is relaxed. In fact, this is a desirable property because it can provide better results, which is the case when comparing the plots in Fig. 6(a) to any of the others.

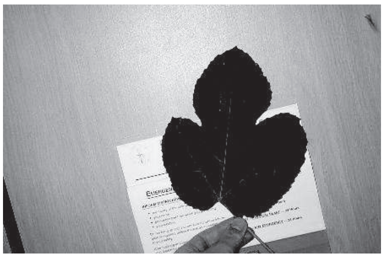


Figure 7: Example showing low NM when compared to Fig. 3(a) using $\mathcal{B} = \{H(f_s), \text{Avg}(f_s)\}$ and $\epsilon = 0.01$

5.2 Supervised learning

Based on the results of the previous section, the NM measure was tested in a supervised learning environment where different sizes of training sets were used as a basis for determining which of the two datasets an unlabelled image came from. In particular, each training set consisted of an equal number of images from both collections. The tests consisted of calculating the NM between the test image and each element in the training set. Then, an average NM was calculated from the measures of each class in the training set. Finally, the unknown image was labelled as coming from the class with the highest average.

SVM were also used to classify the unknown test images. In order to provide a basis for comparison, the same features were used as for the NM. However, a window size of 10×10 would provide vectors with extremely large dimensionality, *e.g.* the dimensionality would be approximately equal to $2 \times M/10 \times N/10$. Thus, in order to avoid the curse of dimensionality (where the number of samples required increases exponentially with the dimensionality of the feature space [27]), the test image was divided into four quadrants and the average grayscale and Pal's entropy were calculated for each quadrant. This produced vectors with eight dimensions for each image. Once the vectors were created, the test images were classified using the Matlab SVM and Kernel Methods Toolbox [41]. The results for both these tests are reported in Table 2. Observe that the NM measure outperforms the SVM for all sizes of training sets.

6 Conclusion

This article presents a practical application of near sets in discovering similar images and in measuring the degree of similarity between images. Near sets themselves reflect human perception, *i.e.*, emulating how humans go about perceiving and, possibly, recognizing objects in the environment. Although a consideration of human perception itself is outside the scope of this article, it should be noted that a rather common sense view of perception underlies the basic understanding of near sets (in effect, perceiving means identifying objects with common descriptions). And perception itself can be understood in Maurice Merleau-Ponty's sense [16], where perceptual objects are those objects captured by the senses. In presenting this application, this article has presented details on how to apply near set theory to the problem of classification of images by way of calculating the nearness of images. The results presented here demonstrate that the NM measure can be used effectively to create pattern classification systems. Moreover, it is the case that the choice of probe functions is very important. The results obtained so far in comparing nearness measures and SVM are promising. Future work in this research includes further comparisons between SVMs and NMs relative to selections of features and corresponding probe functions. For example, it may be that probe functions that are invariant with respect to scale, rotation, and translations would produce closer results. What is certain is that the results presented in this article demonstrate that near set theory can be a useful tool in image recognition systems and that perception based classification is possible.

Acknowledgment

The authors would like to thank Sankar K. Pal, Piotr Wasilewski, Andrzej Skowron, Jarisław Stepaniuk, and Sheela Ramanna for their insights and suggestions concerning topics in this paper. This research has been supported by the Natural Sciences & Engineering Research Council of Canada (NSERC) grant 185986 and a NSERC Postgraduate Fellowship.

References

[1] E. D. Montag and M. D. Fairchild, "Pyschophysical evaluation of gamut mapping techniques using simple rendered images and artificial gamut boundaries," *IEEE Transactions on Image Processing*, vol. 6, no. 7, pp. 997–989, 1997.

[2] I. El-Naqa, Y. Yang, N. P. Galatsanos, R. M. Nishikawa, and M. N. Wernick, "A similarity learning approach to content-based image retrieval: application to digital mammography," *IEEE Transactions on Medical Imaging*, vol. 23, no. 10, pp. 1233–1244, 2004.

[3] M. Rahman, P. Bhattacharya, and B. C. Desai, "A framework for medical image retrieval using machine learning and statistical similarity matching techniques with relevance feedback," *IEEE Transactions on Information Technology in Biomedicine*, vol. 11, no. 1, pp. 58–69, 2007.

[4] J. I. Martinez, A. F. G. Skarmeta, and J. B. Gimeno, "Fuzzy approach to the intelligent management of virtual spaces," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 36, no. 3, pp. 494–508, 2005.

[5] T. V. Papathomas, R. S. Kashi, and A. Gorea, "A human vision based computational model for chromatic texture segregation," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 27, no. 3, pp. 428–440, 1997.

[6] A. Mojsilovic, H. Hu, and E. Soljanin, "Extraction of perceptually important colors and similarity measurement for image matching, retrieval and analysis," *IEEE Transactions on Image Processing*, vol. 11, no. 11, pp. 1238–1248, 2002.

[7] N. Balakrishnan, K. Hariharakrishnan, and D. Schonfeld, "A new image representation algorithm inspired by image submodality models, redundancy reduction, and learning in biological vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 9, pp. 1367–1378, 2005.

[8] A. Qamra, Y. Meng, and E. Y. Chang, "Enhanced perceptual distance functions and indexing for image replica recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 379–391, 2005.

[9] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assesment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[10] L. Dempere-Marco, H. Xiao-Peng, S. L. S. MacDonald, S. M. Ellis, D. M. Hansell, and G.-Z. Yang, "The use of visual search for knowledge gathering in image decision support," *IEEE Transactions on Medical Imaging*, vol. 21, no. 7, pp. 741–754, 2002.

[11] S. Kuo and J. D. Johnson, "Spatial noise shaping based on human visual sensitivity and its application to image coding," *IEEE Transactions on Image Processing*, vol. 11, no. 5, pp. 509–517, 2002.

[12] B. A. Wandell, A. El Gamal, and B. Girod, “Common principles of image acquisition systems and biological vision,” *Proceedings of the IEEE*, vol. 90, no. 1, pp. 5–17, 2002.

[13] T. A. Wilson, S. K. Rogers, and M. Kabrisky, “Perceptual-based image fusion for hyperspectral data,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 4, pp. 1007–1017, 1997.

[14] A. Hoogs, R. Collins, R. Kaucic, and J. Mundy, “A common set of perceptual observables for grouping, figure-ground discrimination, and texture classification,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 4, pp. 458–474, 2003.

[15] N. G. Bourbakis, “Emulating human visual perception for measuring difference in images using an spn graph approach,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 32, no. 2, pp. 191–201, 2002.

[16] M. Merleau-Ponty, *Phenomenology of Perception*. Paris and New York: Smith, Callimard, Paris and Routledge & Kegan Paul, 1945, 1956, trans. by Colin Smith.

[17] D. Calitoiu, B. J. Oommen, and D. Nussbaum, “Desynchronizing a chaotic pattern recognition neural network to model inaccurate perception,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 37, no. 3, pp. 692–704, 2007.

[18] M. Fahle and T. Poggio, *Perceptual Learning*. Cambridge, MA: The MIT Press, 2002.

[19] J. F. Peters, “Classification of objects by means of features,” in *Proceedings of the IEEE Symposium Series on Foundations of Computational Intelligence (IEEE SCCI 2007)*, Honolulu, Hawaii, 2007, pp. 1–8.

[20] ——, “Near sets. general theory about nearness of objects,” *Applied Mathematical Sciences*, vol. 1, no. 53, pp. 2609–2629, 2007.

[21] J. F. Peters, A. Skowron, and J. Stepaniuk, “Nearness of objects: Extension of approximation space model,” *Fundamenta Informaticae*, vol. 79, no. 3-4, pp. 497–512, 2007.

[22] Z. Pawlak, “Classification of objects by means of attributes,” Institute for Computer Science, Polish Academy of Sciences, Tech. Rep. PAS 429, 1981.

[23] ——, “Rough sets,” *International Journal of Computer and Information Sciences*, vol. 11, pp. 341–356, 1982.

[24] Z. Pawlak and A. Skowron, “Rudiments of rough sets,” *Information Sciences*, vol. 177, pp. 3–27, 2007.

[25] ——, “Rough sets: Some extensions,” *Information Sciences*, vol. 177, pp. 28–40, 2007.

[26] ——, “Rough sets and boolean reasoning,” *Information Sciences*, vol. 177, pp. 41–73, 2007.

[27] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd ed. Wiley, 2001.

[28] V. N. Vapnik, *Statistical Learning Theory*. New York: John Wiley & Sons, Inc., 1998.

[29] C. J. C. Burges, “A tutorial on support vector machines for pattern recognition,” *Data Mining and Knowledge Discovery*, vol. 2, pp. 121–167, 1998, http://research.microsoft.com/~cburges/papers/SVMTutorial.pdf.

[30] C. Henry and J. F. Peters, "Image pattern recognition using approximation spaces and near sets," in *Proc. of the Eleventh International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computer (RSFDGrC 2007), Joint Rough Set Symposium (JRS07), Lecture Notes in Artificial Intelligence*, vol. 4482, 2007, pp. 475–482.

[31] ——, "Near set index in an objective image segmentation evaluation framework," in *GEOgraphic Object Based Image Analysis: Pixels, Objects, Intelligence*, University of Calgary, Alberta, 2008, p. to appear.

[32] A. E. Hassanien, A. Abraham, J. F. Peters, G. Schaefer, and C. Henry, "Rough sets and near sets in medical imaging: A review," *IEEE TRansactions on Information Technology in Biomedicine*, p. submitted, 2008.

[33] J. F. Peters, S. Shahfar, S. Ramanna, and T. Szturm, "Biologically-inspired adaptive learning: A near set approach," in *Frontiers in the Convergence of Bioscience and Information Technologies*, Korea, 2007.

[34] J. F. Peters and S. Ramanna, "Feature selection: A near set approach," in *ECML & PKDD Workshop in Mining Complex Data*, Warsaw, 2007, pp. 1–12.

[35] J. F. Peters and P. Wasilewski, "Foundations of near sets," *Elsevier Science*, p. submited, 2008.

[36] J. F. Peters, "Discovery of perceptually near information granules," in *Novel Developements in Granular Computing: Applications of Advanced Human Reasoning and Soft Computation*, J. T. Yao, Ed. Hersey, N.Y., USA: Information Science Reference, 2008, p. to appear.

[37] N. R. Pal and S. K. Pal, "Entropy: A new definition and its applications," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 21, no. 5, pp. 1260 – 1270, 1991.

[38] J. Marti, J. Freixenet, J. Batlle, and A. Casals, "A new approach to outdoor scene description based on learning and top-down segmentation," *Image and Vision Computing*, vol. 19, pp. 1041–1055, 2001.

[39] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proceedings of the 8th International Conference on Computer Visison*, vol. 2, 2001, pp. 416–423.

[40] M. Weber, *Leaves Dataset*. Computational Vision at Caltech, 1999, url: http://www.vision.caltech.edu/archive.html.

[41] S. Canu, Y. Grandvalet, V. Guigue, and A. Rakotomamonjy, "SVM and kernel methods matlab toolbox," 2005.