Signature-based Perceptual Nearness. Application of Near Sets to Image Retrieval

Christopher J. Henry and Sheela Ramanna

Abstract. This paper presents a signature-based approach to quantifying perceptual nearness of images. A signature is defined as a set of descriptors, where each descriptor consists of a real-valued feature vector associated with a digital image region (set of pixels) combined with a region-based weight. Tolerance near sets provide a formal framework for our application of near sets to image retrieval. The tolerance nearness measure tNM was created to demonstrate application of near set theory to the problem of image correspondence. A new form of tNM has been introduced in this work, which takes into account the region size. Our method is compared to two other well-known image similarity measures: Earth Movers Distance (EMD) and Integrated Region Matching (IRM).

MSC2010 Classification: Primary: 54E35 (metric spaces), 62H35 (image analysis), Secondary: 68U10 (image processing), 68N01 (software).

Keywords. Digital image, near sets, perceptual nearness, similarity measure, tolerance.

1. Introduction

Knowledge can be viewed as the theoretical or practical understanding of a subject [1], which can be applied in decision making processes, as in, for example, determining the perceptual nearness of sets of images. Near set theory provides the framework for evaluation of the perceptual similarity of sets of objects [2]. The view of perceptual nearness presented here is rather common sense, where objects, or sets of objects, that *appear* similar are considered perceptually near each other. Here we stress the usage of the term *appear* that relates to impressions [1]. For instance,

our senses (viewed as probe functions [3]) provide details about objects in our environment which we use to describe objects and discern patterns of interest. Decisions on nearness are then made within the context of a perceptual system, *i.e.* a system consisting of objects and our impressions of these objects. This form of perception follows from work by Zeeman [4, 5] (with respect to visual perception) and Riesz (with respect to proximity) [6, 7], which has spawned many recent publications [8, 9, 10, 11, 12].

Depending on experience and training, as well as, the application, knowledge can be used to make specific decisions based on relevant features, which are gleaned from our senses. For example, a doctor looking at MR images will base a decision within a different perceptual system than a photographer assessing the similarity of natural images. In these cases, the expert is trained to perceive specific characteristics in order to assess similarity and make an informed decisions. Thus, knowledge, or the application of knowledge, is employed within perceptual systems.

The approach presented here models the application of knowledge based on sets of objects with similar descriptions (features) within the framework of tolerance near set theory [13, 14, 15]. Objects are grouped by way of finding tolerance classes, *i.e.* classes where the difference in descriptions of each pair of objects from the set is less than some ε. By extracting and collecting the data in this manner, assessing the similarity of objects and sets of objects in terms of high-level concepts is possible. Notice, this approach starts with the selection of features (within a perceptual system), which defines the context for understanding a subject. Thus, tolerance near sets provide a formal framework for the application of knowledge in assessing perceptual nearness.

This paper presents a signature-based approach to quantifying perceptual nearness of images and is an extension of work reported in [16]. A signature is defined as a set of descriptors, where each descriptor consists of a real-valued feature vector associated with a digital image region (set of pixels) combined with a region-based weight. Tolerance near sets provide a formal framework for our application of near sets to image retrieval [17]. The tolerance nearness measure tNM was created to demonstrate application of near set theory to the problem of image correspondence (see, *e.g.* [11, 18, 16]). Our method is compared to two other well-known image similarity measures: Earth Movers Distance (EMD) [19] and Integrated Region Matching (IRM) [20, 21]. The contribution of this article is a new form of the tNM which takes into account the region size, for use in signature-based applications.

The paper is organized as follows. First, an introduction to near set theory is given in Section 2. Next, Sections 3 - 6 introduce the signature-based methods used to generate results. Implementation details are reported in Section 7, and results and discussion are given in Section 8.

2. Background: Near Set Theory

A logical starting point for a discussion on near set theory begins with establishing a basis for describing elements of sets. All sets in near set theory consist of perceptual objects. A *perceptual object* is something that has its origin in the physical world. A perceptual object is anything in the physical world with characteristics observable to the senses such that they can be measured and are knowable to the mind. In keeping with the approach to pattern recognition suggested by M. Pavel [3], the features of a perceptual object are quantified by probe functions, where a *feature* characterizes some aspect of the makeup of a perceptual object [22]. A *probe function* is a real-valued function representing a feature of a perceptual object [23, 24]. Next, a perceptual system is a set of perceptual objects, together with a set of probe functions. Formally, a *perceptual system* $\langle O, \mathbb{F} \rangle$ consists of a non-empty set O of sample perceptual objects and a non-empty set $\mathbb{F}$ of real-valued functions $\phi \in \mathbb{F}$ such that $\phi : O \to \mathbb{R}$ [2].

Combining the concepts of a perceptual object and a probe function, the description of a perceptual object within a perceptual system can be defined as follows. Let $\langle O, \mathbb{F} \rangle$ be a perceptual system, and let $\mathcal{B} \subseteq \mathbb{F}$ be a set of probe functions. Then, the *description* of a perceptual object $x \in O$ is a feature vector given by

$$\Phi_{\mathcal{B}}(x) = (\phi_1(x), \phi_2(x), \ldots, \phi_i(x), \ldots, \phi_l(x)),$$

where l is the length of the vector $\Phi_{\mathcal{B}}$, and each $\phi_i(x)$ in $\Phi_{\mathcal{B}}(x)$ is a probe function value that is part of the description of the object $x \in O$. Note, the idea of a feature space is implicitly introduced along with the definition of object description. An object description is the same as a feature vector as described in traditional pattern classification [25]. The description of an object can be considered a point in an l-dimensional Euclidean space $\mathbb{R}^l$ called a feature space. Here, it is important to note that near set theory is concerned with the nearness of objects based on their descriptions. Thus, the relationship between objects is discovered in a feature space that is determined by the probe functions in $\mathcal{B}$.

In this work, the relation between objects is defined within the context of a tolerance space [15], which is required for real world problems and applications [26, 27].

Definition 1. Tolerance Space [4, 26]. *Let O be a set of sample perceptual objects, and let ξ be a binary relation (called a tolerance relation) on $X \subseteq O$ ($\xi \subset X \times X$) that is reflexive (for all $x \in X$, $x\xi x$) and symmetric (for all $x, y \in X$, if $x\xi y$, then $y\xi x$) but transitivity of ξ is not required. Then a tolerance space is defined as $\langle X, \xi \rangle$.*

Next, a specific tolerance relation is given in Definition 2.

Definition 2. Perceptual Tolerance Relation [13, 14] (see [28, 29] for applications). *Let $\langle O, \mathbb{F} \rangle$ be a perceptual system and let $\varepsilon \in \mathbb{R}$. For every $\mathcal{B} \subseteq \mathbb{F}$, the* perceptual tolerance relation $\cong_{\mathcal{B},\varepsilon}$ *is defined as follows:*

$$\cong_{\mathcal{B},\varepsilon} = \{(x, y) \in O \times O :\ \parallel \Phi(x) - \Phi(y) \parallel_2\ \leq \varepsilon\},$$

where $\parallel \cdot \parallel_2$ is the L^2 norm. For notational convenience, this relation is written $\cong_{\mathcal{B}}$ instead of $\cong_{\mathcal{B},\varepsilon}$ with the understanding that ε is inherent to the definition of the tolerance relation.

The following examples highlights the effect of the perceptual tolerance relation. First, consider the example given in Fig. 1, where the oval in Fig. 1a represents a set of objects, and objects are grouped together into classes (represented by the different colours) such that each pair of objects in the class satisfies the tolerance relation. Moreover, in this figure, the position of a class is also relevant. Here, the distance between the object descriptions in feature space increases with the distance between classes in the image. Starting on the left side, Fig. 1a represents the case when $\varepsilon = 0$. Thus each class contains objects with matching descriptions (called an equivalence class). In accordance, Fig. 1b & 1c represents the classes that result from a low and high value of epsilon, which respectively correspond to the object descriptions of two adjacent equivalence classes satisfying Definition 2, and the object descriptions of four adjacent classes satisfying Definition 2. Observe that low values of epsilon tend to produce a large number of small classes, and high value of epsilon tend to produce a small number of large classes.

The next example demonstrates the effect of the perceptual tolerance relation on real data. Consider Table 1 that contains 20 objects with $l = 1$ (*i.e.* the object description is a
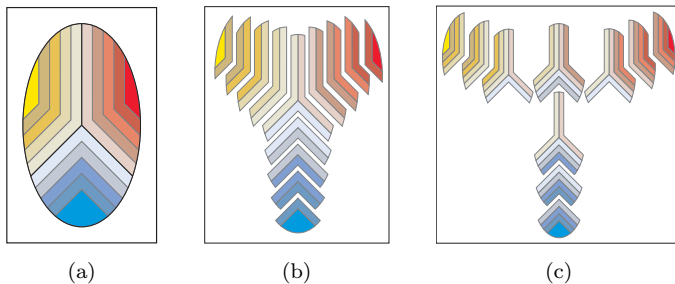
(a)   (b)   (c)

FIGURE 1. Example highlighting the effect of the perceptual tolerance relation.

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

vector of length 1). Letting $\varepsilon = 0.1$ gives the following classes:

$$\{\{x_1, x_8, x_{10}, x_{11}\}, \{x_1, x_9, x_{10}, x_{11}, x_{14}\},$$
$$\{x_2, x_7, x_{18}, x_{19}\},$$
$$\{x_3, x_{12}, x_{17}\},$$
$$\{x_4, x_{13}, x_{20}\}, \{x_4, x_{18}\},$$
$$\{x_5, x_6, x_{15}, x_{16}\}, \{x_5, x_6, x_{15}, x_{20}\},$$
$$\{x_6, x_{13}, x_{20}\}\}$$

Observe that each pair of objects in each of the above classes satisfies the condition $\| \phi(x) - \phi(y) \|_2 \leq \varepsilon$, and that almost all of the objects appear in more than one class. Moreover, there would be twenty classes, if the perceptual indiscernibility relation were used since there are no two objects with matching descriptions.

In each of the previous examples, the objects are grouped together based on the following definitions.

Definition 3. Pre-Class [30]. *Let $\langle O, \mathbb{F} \rangle$ be a perceptual system. For $\mathcal{B} \subseteq \mathbb{F}$ and $\varepsilon \in \mathbb{R}$, a set $X \subseteq O$ is a* pre-class *iff $x \cong_{\mathcal{B}, \varepsilon} y$ for any pair $x, y \in X$.*

Definition 4. Tolerance Class [30]. *A maximal pre-class with respect to inclusion is called a tolerance class.*

Observe that each pair of objects must satisfy the tolerance relation in a pre-class, *i.e.*, the distance (using the L^2 norm) between each pair of objects in a pre-class is less than or equal to ε. Also, objects can belong to more than one pre-class, and the only difference between Definitions 3 & 4 is that for any given tolerance class, C, no other objects in O can be added to it.

Since objects can belong to more than one tolerance class the following notation is required to differentiate between classes and facilitate discussions in subsequent sections. The set of all tolerance classes using only the objects in O is given by $H_{\cong_{\mathcal{B}, \varepsilon}}(O)$ (also called the cover of O), a single tolerance class is represented by $C \in H_{\cong_{\mathcal{B}, \varepsilon}}(O)$, and the set of all tolerance classes containing an object x is denoted by $C_x \subset H_{\cong_{\mathcal{B}, \varepsilon}}(O)$.

Finally, this section is concluded by defining near sets. Recall, the idea behind near set theory is that sets of objects that have similar descriptions are considered near each other. Furthermore, a method for determining similarity was provided by way of the perceptual tolerance relation. Consequently, the following two definitions enunciate the fundamental notion of nearness between two sets and provide the foundation of the results presented in this article.

Definition 5. Tolerance Nearness Relation [13, 14]. *Let $\langle O, \mathbb{F} \rangle$ be a perceptual system and let $X, Y \subseteq O, \varepsilon \in \mathbb{R}$. A set X is near to a set Y within the perceptual system $\langle O, \mathbb{F} \rangle$ $(X \bowtie_{\mathbb{F}} Y)$ iff there exists $x \in X$ and $y \in Y$ and there is $\mathcal{B} \subseteq \mathbb{F}$ such that $x \cong_{\mathcal{B}, \varepsilon} y$. Sets X and Y are tolerance near sets iff X is near to Y.*

3. Signature-based Measures

This section presents the signature-based methods investigated in this article. Generally, signatures are mathematical descriptions of an image [31], where specific signature details are dependent on the application. A signature is defined as a set of pairs, where each pair consists of a feature vector associated with a set of pixels (region) from an image combined with the cardinality of the set. We obtained our signatures by first segmenting an image, *i.e.* partitioning an image into non-overlapping regions (described in Section 7). Then, image colour,

edge, and texture features were used to generate the final signature (again, see Section 7). The following gives the formal definition of a signature (inspired by the definition in [19]). First, define an RGB image as $F = \{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_T\}$, where $\mathbf{p}_i = (c, r, R, G, B)^{\mathrm{T}}$, $c \in [1, M]$, $r \in [1, N]$, $R, G, B \in [0, 255]$, and M, N respectively denote the width and height of the image and $M \times N = T$. Further, define a region (also known as an image segment) of F as $f_i \subseteq F$ such that $f_i \cap f_j = \emptyset$ for $i \neq j$ and $f_1 \cup f_2 \cdots \cup f_s = F$, where s is the number of regions in the partition of F. Next, a specific region descriptor can be defined as $s_j^F = (\mathbf{r}_j^F, w_{\mathbf{r}_j}^F)$, where F denotes the image from which the descriptor belongs (necessary for the methods introduced below), $\mathbf{r}_j^F$ is a real-valued feature vector, $w_{\mathbf{r}_j}^F = |f_j|/|F|$, and $0 \leq j < s$. Note, $\mathbf{r}_j^F$ is called a cluster representative in [19], the region descriptor in [21], and, in the context of the signature-based tolerance nearness measure presented below, $\Phi_{\mathcal{B}}(s_j^F) = \mathbf{r}_j^F$, where each specific value of $\mathbf{r}_j^F$ is obtained by some $\phi_i(s_j^F)$. Finally, the set of all descriptors extracted from an image F is called the image signature and is defined as $S_F = \{s_j^F\}$.

4. Earth Mover's Distance

The EMD was introduced by Rubner in [19] and is also known as Mallows distance when applied to probability frequencies [31]. The EMD is based on the idea of minimizing the amount of work required to move multiple piles of dirt to a series of holes in the ground. In terms of measuring image similarity the piles of dirt and holes are represented by image signatures, where the location of the dirt piles (resp. holes) in feature space is determined by the feature vector and the size of the pile (hole) is determined by the region count. The EMD is calculated by solving the transportation problem [32], subject to constraints, where signatures from the respective images are cast as consumers and suppliers.

Specifically, for two images P and Q (with S_P and S_Q), let $d(\mathbf{r}_i^P, \mathbf{r}_j^Q)$ be the distance between region cluster representatives. Then, to calculate the EMD, it is necessary to find a flow matrix $F = [f_{i,j} : i = 1, \ldots, s_P; j = 1, \ldots, s_Q]$ that minimizes

$$\mathrm{WORK}(S_P, S_Q, F) = \sum_{i=1}^{s_P} \sum_{j=1}^{s_Q} d(\mathbf{r}_i^P, \mathbf{r}_j^Q) f_{i,j},$$

subject to the constraints (reported in [19]) in Eq. 4.1 - 4.4, where each constraint is described as follows. Eq. 4.1 ensures that the "supples" only move from P to Q, Eq. 4.2 restricts a cluster in P to only transmitting the amount of supplies specified by its weight, Eq. 4.3 limits clusters

in Q to receiving supplies amounting to its weight, and Eq. 4.4 requires the maximum of amount of supplies to be transported.

$$f_{i,j} \geq 0 \quad 1 \leq i \leq s_P, 1 \leq j \leq s_Q \quad (4.1) \qquad \sum_{j=1}^{s_Q} f_{i,j} \leq w_{\mathbf{r}_i}^P \quad 1 \leq i \leq s_P \qquad (4.3)$$

$$\sum_{i=1}^{s_P} f_{i,j} \leq w_{\mathbf{r}_j}^Q \quad 1 \leq j \leq s_Q \qquad (4.2) \qquad \sum_{i=1}^{s_P} \sum_{j=1}^{s_Q} f_{i,j} = \min\left(\sum_{i=1}^{s_P} w_{\mathbf{r}_i}^P, \sum_{j=1}^{s_Q} w_{\mathbf{r}_j}^Q \right) \quad (4.4)$$

To calculate the flow, we used the code available at [33] (reported in [34, 19]) that uses the transportation-simplex method [35]. Once the optimal flow F has been calculated, the EMD is defined as

$$\mathrm{EMD}(S_P, S_Q) = \frac{\sum_{i=1}^{s_P} \sum_{j=1}^{s_Q} d(\mathbf{r}_i^P, \mathbf{r}_j^Q) f_{i,j}}{\sum_{i=1}^{s_P} \sum_{j=1}^{s_Q} f_{i,j}}.$$

5. Integrated Region Matching Similarity Measure

The IRM similarity measure is a soft matching approach to measuring the similarity of images [20, 21]. Soft matching techniques allow multiple matches between segments to reduce the effect of segmentations that do not match our perception of the objects in the images (*i.e* to reduce the effect of poor image segmentations) [31]. Here, it is important to differentiate between segments or regions, and perceptual concepts within an image. Let us define regions as the output of an image segmentation algorithm designed to isolate perceptual concepts (*i.e.* areas containing specific perceptual or semantic meaning) within the image. In other words, due to improper segmentation, it is possible to have multiple segments or regions per perceptual concept. Using these definitions, it is easy see the advantage of associating more than one segment with a region, as is the case in matching the multiple brown horse segments in Fig. 2c with the single brown segment in Fig. 2d.

The IRM similarity measure is calculated by a weighted sum of the distance between region feature vectors, where weights are determined by a significance matrix containing the significance of matching regions in the two respective images. The significance matrix is populated by an algorithm that attempts to assign the highest value of significance to regions that are the most similar, where similarity is defined with respect to distance between region feature vectors and region size.

Formally, the IRM similarity measure is calculated as follows [21]. Let $S = [s_{i,j} : i = 1, \ldots, s_P; j = 1, \ldots, s_Q]$ represent a significance matrix indicating the importance between

(a)   (b)   (c)   (d)
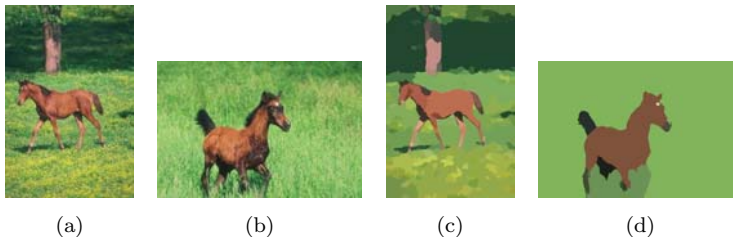
FIGURE 2. Multiple segments per region example.

matching region descriptor $\mathbf{r}_i^P$ with region descriptor $\mathbf{r}_j^Q$. Then, the IRM similarity measure is defined as

$$\text{IRM}(S_P, S_Q) = \sum_{i=1}^{s_P} \sum_{j=1}^{s_Q} s_{i,j} d(\mathbf{r}_i^P, \mathbf{r}_j^Q).$$

Notice, the key to calculating the IRM similarity measure is in populating the significance matrix. As was the case for the EMD, this task is performed subject to the constraints. Namely, all regions must play a role for measuring similarity and the most similar regions must be assigned the highest priority (see, *e.g.*, [21]). The algorithm used to populate S is given in [21].

6. Tolerance Near Set Nearness Measure

The tolerance nearness measure was created out of a need to determine the degree that near sets resemble each other, a need which arose during the application of near set theory to the practical applications of image correspondence (see, *e.g.* [28, 11]). The tolerance nearness measure between two sets X, Y is based on the idea that tolerance classes formed from objects in the union $Z = X \cup Y$ should be evenly divided among X and Y if these sets are similar, where similarity is always determined with respect to the selected probe functions*. The tolerance nearness measure between two sets X and Y, is defined as follows.

Definition 6. Tolerance Nearness Measure. *Let $\langle O, \mathbb{F} \rangle$ be a perceptual system, with $\varepsilon \in \mathbb{R}$, and $\mathcal{B} \subseteq \mathbb{F}$. Furthermore, let $X, Y \subseteq O$ be two disjoint sets and let $Z = X \cup Y$. Then, a tolerance*

---

*Perceptual information is always presented with respect to the probe functions contained in $\mathcal{B}$ just as our senses define our perception of the world. For example our ability to view light in the visible spectrum rather than infra red or microwaves spectra defines our perception of the world just as the selection of probe functions constrains the amount of perceptual information available for extraction from a set of objects.

nearness measure between X and Y is given by

$$tNM_{\cong_{\mathcal{B},\varepsilon}}(X,Y) =$$

$$1 - \left(\sum_{C \in H_{\cong_{\mathcal{B},\varepsilon}}(Z)} W \right)^{-1} \cdot \sum_{C \in H_{\cong_{\mathcal{B},\varepsilon}}(Z)} W \frac{\min(|C \cap X|, |[C \cap Y|)}{\max(|C \cap X|, |C \cap Y|)}. \quad (6.1)$$

Traditionally, the tNM has been used in measuring nearness in problem domains that generate many objects for comparison. For example, in [11], images are divided into subimages, where each subimage is an object. Keeping the subimage size relative small (with respect to the size of the image) creates many objects for comparison. Thus, in Eq 6.1, the weight W is usually defined as cardinality of the tolerance class C (*i.e* $W = |C|$). In terms of signatures, it is conceivable that there are many of these subimages per region, and in fact this is usually the case. However, in the comparison of image signatures, tolerance classes will likely be small since there will be either two specific region descriptors for two similar regions from the respective images, or small groups of region descriptors from each region (as depicted in Fig. 2).

In order to provide a basis for comparison, the tolerance near set approach was adapted to take into consideration the fact that we are matching regions, with few or even one signature, rather than a large set of objects. In particular, to calculate the tolerance nearness measure between two images P and Q, the objects in the perceptual system are signatures derived from P and Q, the object descriptions are given by $\Phi_{\mathcal{B}}(s_j^F) = \mathbf{r}_j^F$, and $X = S_P$ and $Y = S_Q$. As a result, each $C \in H_{\cong_{\mathcal{B},\varepsilon}}(O)$ is a set of signatures, and the weight in Eq. 6.1 is now defined as

$$W = \sum_i w_{\mathbf{r}_i}^P + \sum_j w_{\mathbf{r}_j}^Q, \quad (6.2)$$

for all $s_j^P, s_j^Q \subseteq C$.

The following example demonstrates calculating the tNM on signatures extracted from the two grayscale images given in Fig. 3. Let $\mathcal{B} \subseteq \mathbb{F}$ contain a single probe function that returns the grayscale value of a region (*i.e.* a value in the interval $[0, 255]$). Next, recall that a region descriptor is defined as $s_j^F = (\mathbf{r}_j^F, w_{\mathbf{r}_j}^F)$, where F denotes the image from which the descriptor belongs, $\mathbf{r}_j^F$ is a real-valued feature vector (in this case a vector of length 1), and $w_{\mathbf{r}_j}^F$ represents the number of pixels contained in the region. Then, signatures from images P and Q are given

as

$$S_P = \{(\mathbf{r}_1^P, w_{\mathbf{r}_1}^P), (\mathbf{r}_2^P, w_{\mathbf{r}_2}^P), (\mathbf{r}_3^P, w_{\mathbf{r}_3}^P), (\mathbf{r}_4^P, w_{\mathbf{r}_4}^P)\},$$

$$= \{(190, 100), (210, 50), (230, 50), (90, 200)\},$$

$$S_Q = \{(\mathbf{r}_1^Q, w_{\mathbf{r}_1}^Q), (\mathbf{r}_2^Q, w_{\mathbf{r}_2}^Q), (\mathbf{r}_3^Q, w_{\mathbf{r}_3}^Q), (\mathbf{r}_4^Q, w_{\mathbf{r}_4}^Q), (\mathbf{r}_5^Q, w_{\mathbf{r}_5}^Q), (\mathbf{r}_6^Q, w_{\mathbf{r}_6}^Q), (\mathbf{r}_7^Q, w_{\mathbf{r}_7}^Q)\}$$

$$= \{(170, 50), (190, 50), (210, 50), (230, 50), (0, 100), (100, 50), (70, 50)\}.$$

Moreover, let $\varepsilon = 0.234$, *i.e.*, only grayscale values that are within 60 of each other satisfy the tolerance relation. Then, $Z = S_P \cup S_Q$, and

$$H_{\cong_{\mathcal{B},\varepsilon}}(Z) = \{\{(\mathbf{r}_1^P, w_{\mathbf{r}_1}^P), (\mathbf{r}_2^P, w_{\mathbf{r}_2}^P)\}, (\mathbf{r}_3^P, w_{\mathbf{r}_3}^P), (\mathbf{r}_1^Q, w_{\mathbf{r}_1}^Q), (\mathbf{r}_2^Q, w_{\mathbf{r}_2}^Q), (\mathbf{r}_3^Q, w_{\mathbf{r}_3}^Q), (\mathbf{r}_4^Q, w_{\mathbf{r}_4}^Q)\},$$

$$\{(\mathbf{r}_4^P, w_{\mathbf{r}_4}^P), (\mathbf{r}_6^Q, w_{\mathbf{r}_6}^Q), (\mathbf{r}_7^Q, w_{\mathbf{r}_7}^Q)\},$$

$$\{(\mathbf{r}_5^Q, w_{\mathbf{r}_5}^Q)\}\}.$$

Finally, the nearness measure is calculated as

$$tNM_{\cong_{\mathcal{B},\varepsilon}}(X, Y) = \frac{1}{800}\left(400 \cdot \frac{3}{4} + 300 \cdot \frac{1}{2} + 100 \cdot \frac{0}{1}\right) = \frac{9}{16}.$$

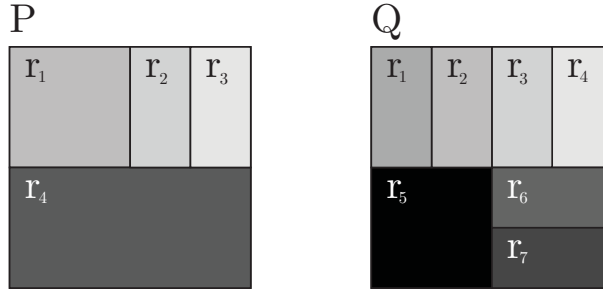FIGURE 3. Sample images used for demonstrating calculation of tNM.

As was mentioned, the main difference between the tNM defined here and the one reported in [11], is the definition of W in Eq. 6.2. In [11], the cardinality of C is used to weight the importance of the fraction in Eq. 6.1 (*i.e.* $W = |C|$) due to the intuitive idea that there should be many subimages per region. In contrast, Eq. 6.2 weights the importance of the fraction by the number of pixels (the size of the regions) in C. Here, the idea is that similar images should have signatures with similar region counts. In addition, region size is also used as an additional feature in finding tolerance classes in order to better compete with the EMD and

IRM similarity measure. In both of the latter approaches, region size plays a significant role in determining similarity. Consequently, the results presented here were actually generated using $\phi_{\mathcal{B}}(s_j^F) = (\mathbf{r}_j^F, w_{\mathbf{r}_j}^F)$, *i.e.*, the size of a region was also used as part of the object description in finding tolerance classes. Note, we did not consider adding this information when generating results with the EMD and IRM similarity measure due to the important role this information places in both the algorithms.

Finally, this section ends by presenting some properties of tNM. The nearness measure produces values in the interval $[0, 1]$, where, for a pair of sets X, Y, a value of 1 represents no resemblance in terms of the probe functions in $\mathcal{B}$, and a value of 0 indicates the sets X, Y completely resemble each other, a fact that can be seen by calculating the nearness measure on a single set, *i.e.* $tNM_{\cong_{\mathcal{B},\varepsilon}}(X, X) = 0$. In addition, the nearness measure provides a methodical approach for determining whether two sets are tolerance near sets as seen in the following proposition and its corollary.

Proposition 1. *A nearness measure of $tNM_{\cong_{\mathcal{B},\varepsilon}}(X, Y) = 1$ indicates that the sets X, Y are not tolerance near sets.*

Proof. A $tNM_{\cong_{\mathcal{B},\varepsilon}}(X, Y) = 1$ is produced by either the empty set (*i.e.* $Z = X \cup Y = \emptyset$) or each tolerance class in $H_{\cong_{\mathcal{B},\varepsilon}}(Z)$ is either completely a subset of X or completely a subset of Y. In other words, there is no tolerance class that consists of objects from both X and Y. Consequently, there is no $x \in X$ and $y \in Y$ such that $x \cong_{\mathcal{B},\varepsilon} y$, and so, by Definition 5, the sets X and Y are not tolerance near sets. $\square$

Corollary 1. *A nearness measure of $tNM_{\cong_{\mathcal{B},\varepsilon}}(X, Y) < 1$ indicates that the sets X, Y are tolerance near sets.*

Proof. A $tNM_{\cong_{\mathcal{B},\varepsilon}}(X, Y) < 1$ is produced, if there is at least one tolerance class $C \in H_{\cong_{\mathcal{B},\varepsilon}}(Z)$ such that $X \cap C \neq \emptyset$ and $Y \cap C \neq \emptyset$. Consequently, there must be $x \in X$ and $y \in Y$ that satisfies $x \cong_{\mathcal{B},\varepsilon} y$, and so, by Definition 5, the sets X and Y are tolerance near sets. $\square$

7. Implementation

All the signatures used to generate our results were extracted from image segments created by an adaptive mean shift segmentation algorithm. The mean shift algorithm, introduced in [36], creates segments based on the assumption that the image can be represented by a mixture model of multivariate density functions. For each pixel, the mean shift algorithm iteratively

searches for a mode (peak) in the local density. Then, a pixel is assigned to the region for which all pixels have the same mode (peak) [37]. The process of finding the modes for an image is based on kernel density estimation, which is a nonparametric technique for estimating the probability density function of a random variable based on observations. Specifically, both the number of observations within a volume in d-dimensional space and a kernel that weights the importance of the observations determines estimate of the distribution [25]. The mean shift segmentations used in this article were created using our own modification of the EDISON system [38], a system for which both the source code and binaries are freely available on line.

The main disadvantage of the mean shift algorithm is the process of selecting the input bandwidth parameter, which defines the geometry of the d-dimensional volume used to select observations for calculation of the density estimation [39]. Selecting an optimal global bandwidth parameter for databases of varying image content is unlikely, and manual selection of the bandwidth parameter for each image is unpractical. A solution to this problem is to select the bandwidth parameter based on the image being segmented. Consequently, the approach reported in [40] (also described in [25]) was used, where, for each pixel in the image, the bandwidth parameter is the distance to its kth nearest neighbourhood. Briefly, given a series of points in $\mathbb{R}^d$, the k-nearest neighbour search problem consists of finding the k-nearest neighbours to a query point q using a specific distance[†].

Kernel density estimators that vary the bandwidth parameter based on the kth-nearest neighbour are called balloon density estimators [39]. The idea is to use a small bandwidth in tightly clustered regions, and a large value in sparse regions. Intuitively, this can be achieved by setting the bandwidth parameter as the distance to the kth-nearest neighbour. Clearly, the success of this approach relies on a fast solution to the k-nearest neighbour search problem. To solve this problem the KNN CUBLAS GPU implementation reported in [41] was used. Note, there are some disadvantages of using the balloon density approach. However, our aim was to either relieve the burden of finding a globally acceptable bandwidth or having to select a bandwidth parameter for each image. The kth-nearest neighbour solved both these problems, while providing good segmentations (see, *e.g.*, Fig. 4).

7.1. Signature Feature Values

The signatures used to generate the results in this paper were based on colour, texture, and edge based features, which is common in content-based image retrieval applications [42, 43, 44,

---

[†]The Euclidean distance was used to produce the results in this article.

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

FIGURE 4. Adaptive mean shift samples.

45]. Starting with edges, this information is represented both implicitly and explicitly in each signature. First, edge information is used indirectly by considering that each image is partitioned into segments whose boundaries, in general, fall along the edges within the image (see, *e.g.* Fig. 4). Next, the edge information is used directly in calculating the texture-based feature. Specifically, the range value (maximum - minimum) for each 3×3 pixel neighbourhood was used as the texture feature to generate the results. Unfortunately, this approach also highlights the edges in images. As a result, the output of a Canny edge detection algorithm was dilated (using morphological image processing techniques) and subtracted from the texture feature output to remove redundant data inherently contained in the image segments. In terms of colour, the three component colours from the CIELAB colour model were used as the colour features[‡]. Examples of each feature can be see in Fig. 5, where Fig. 5a is the original image, Fig 5b is the range value of each 3×3 pixel neighbourhood, Fig. 5c is the result of the Canny edge detection algorithm, Fig. 5d is the morphological dilation of Fig. 5c using a diamond shaped structuring element, Fig. 5e is the difference between Fig. 5b and Fig. 5d, and Fig. 5f - 5h are the CIELAB colour components. Finally, the actual region descriptor consisted of the texture feature depicted in Fig. 5e, and the three LAB colour values. Note, each value is an average over the entire segment.

8. Results and Discussion

The choice of application for comparison of the signature-based methods presented here is in the area of content-based image retrieval (see, *e.g.*, [42]), where the goal is to retrieve images from databases based on content of an image rather than on some semantic string or keywords associated with the image. Specifically, the CBIR results were generated with a 1000 image subset of the COREL dataset [46]. These 1000 images were divided into 10 categories and the

---

[‡]LAB values are obtained from RGB using a white point fixed at (0.9642, 1, 0.8249), which is the D65 standard.
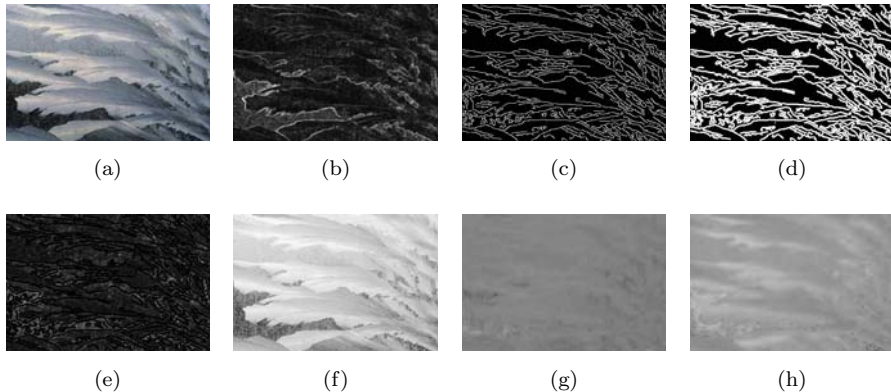
FIGURE 5. Example showing extraction of signature feature values. (a) Original image, (b) the range value of each 3×3 pixel neighbourhood, (c) result of the Canny edge detection algorithm, (d) morphological dilation of (c), (e) difference image between (b) and (d), and (f)-(h) are the CIELAB colour components.

image dimensions were either 384×256 or *vice versa*. Moreover, the categories are varied with different objects and scenes, and images in different categories can also resemble each other (see, *e.g.* Fig. 6).

The results are presented in Fig. 7 & 8. In Fig. 7, each plot represents best query for each category, where the best result is defined as the query image that generated precision vs. recall plots with the most area under the curve. Similarly, the results presented in Fig. 8 represent the average precision vs. recall values over the 100 images in each category. Starting with IRM, notice this measure tends to perform the worst, which may be attributed to the fact that this method was designed to measure the similarity of images with at most 16 regions (and, hence, 16 signatures). In our case, the number of regions used to generate our results were in the interval $[25, 1024]$. Next, Fig. 8 indicates the tNM performs better in categories 1, 3, 5, and 8, and performs similarly to the EMD in categories 0, 2, and 9. However, of these latter categories, the best query for tNM initially returns more correctly retrieved images than the EMD in categories 0, and 2, which is indicated by a large number of precision values at 100%. Similarly, in the case of category 6 the tNM correctly retrieves more images from the same category before a false positive occurs. Lastly, the EMD tends to do better in the categories that
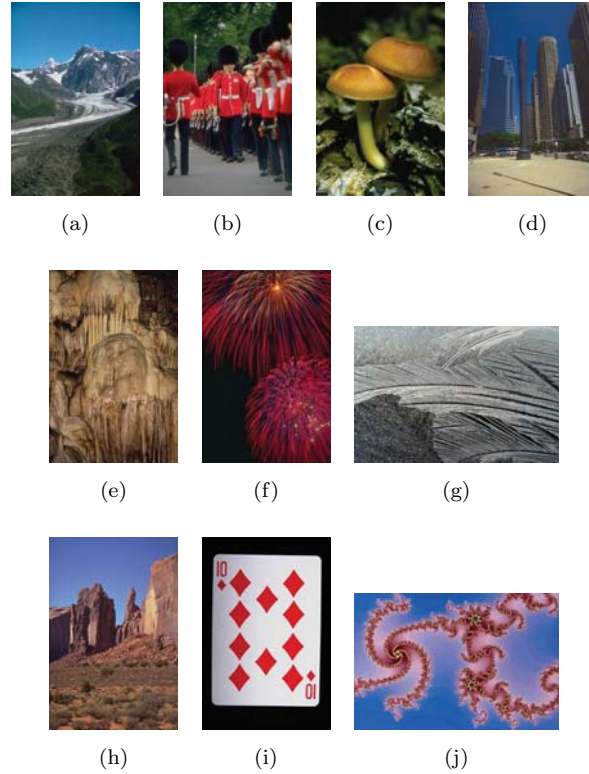
FIGURE 6. Examples of each category of images. (a) - (j) Categories 0 - 9.

contain mostly texture images (*i.e.* categories 4 and 6), which suggests that the EMD approach is better able resolve similarities using the selected texture feature than the tNM approach.

Next, a few comments on the quality of the precision vs. recall values. The ideal case occurs when all the images from the same category as the query image are retrieved before any other images. In this case, precision will be 100% until recall reaches 100%. As can be seen in Fig. 7, the results are quite removed from the ideal case. This is due to the fact that the selected test set reflects a real world database where images in all categories are quite similar. This is not to say that the measures are returning disimiar images, only that the similarity of images (using the selected probe fucntions) does agree with the defined categories. However, this test set was adequate in establishing a benchmark that indicates that the proposed method is, in some instances, better than the established IRM similairty measure and the EMD.
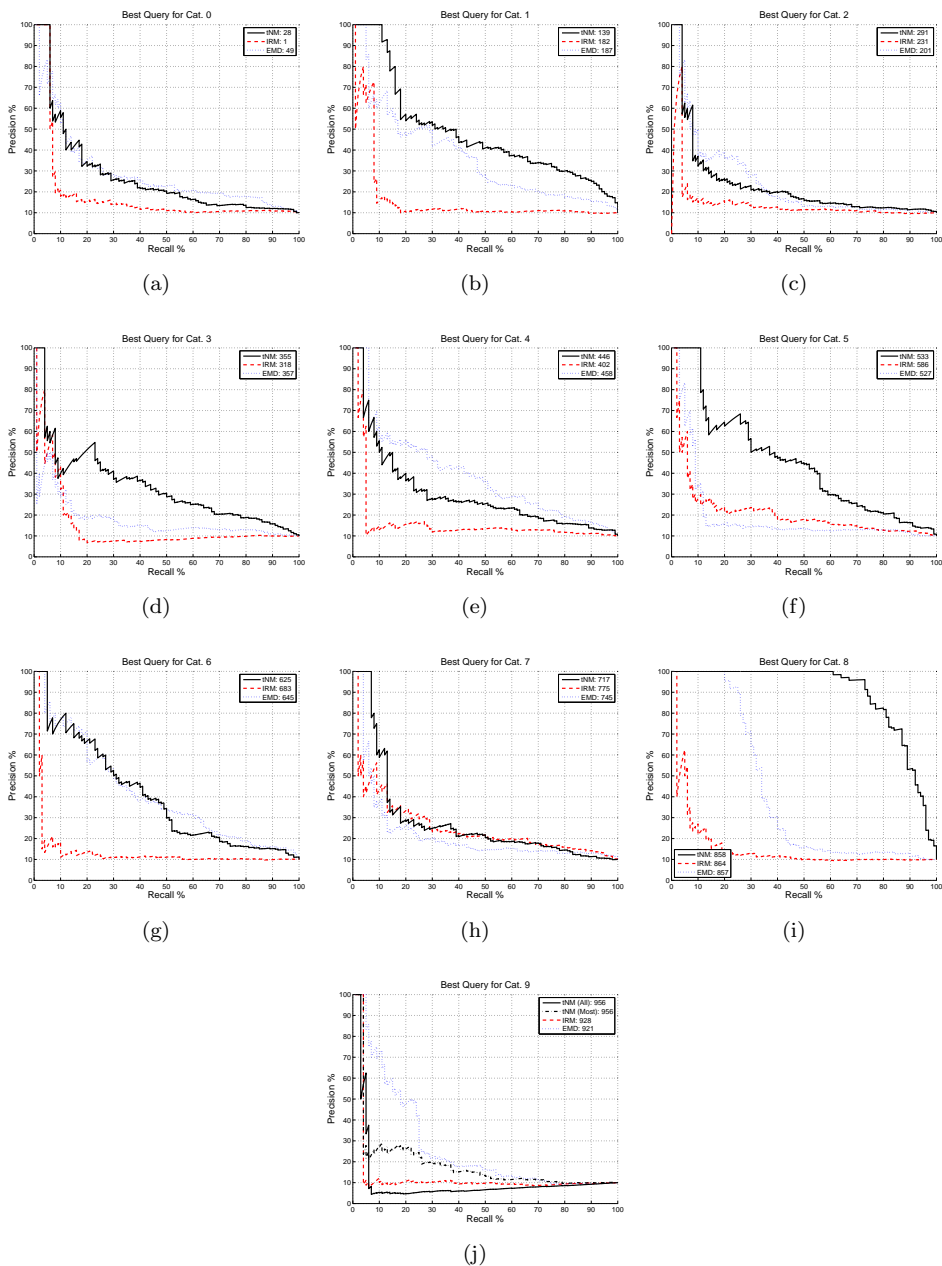
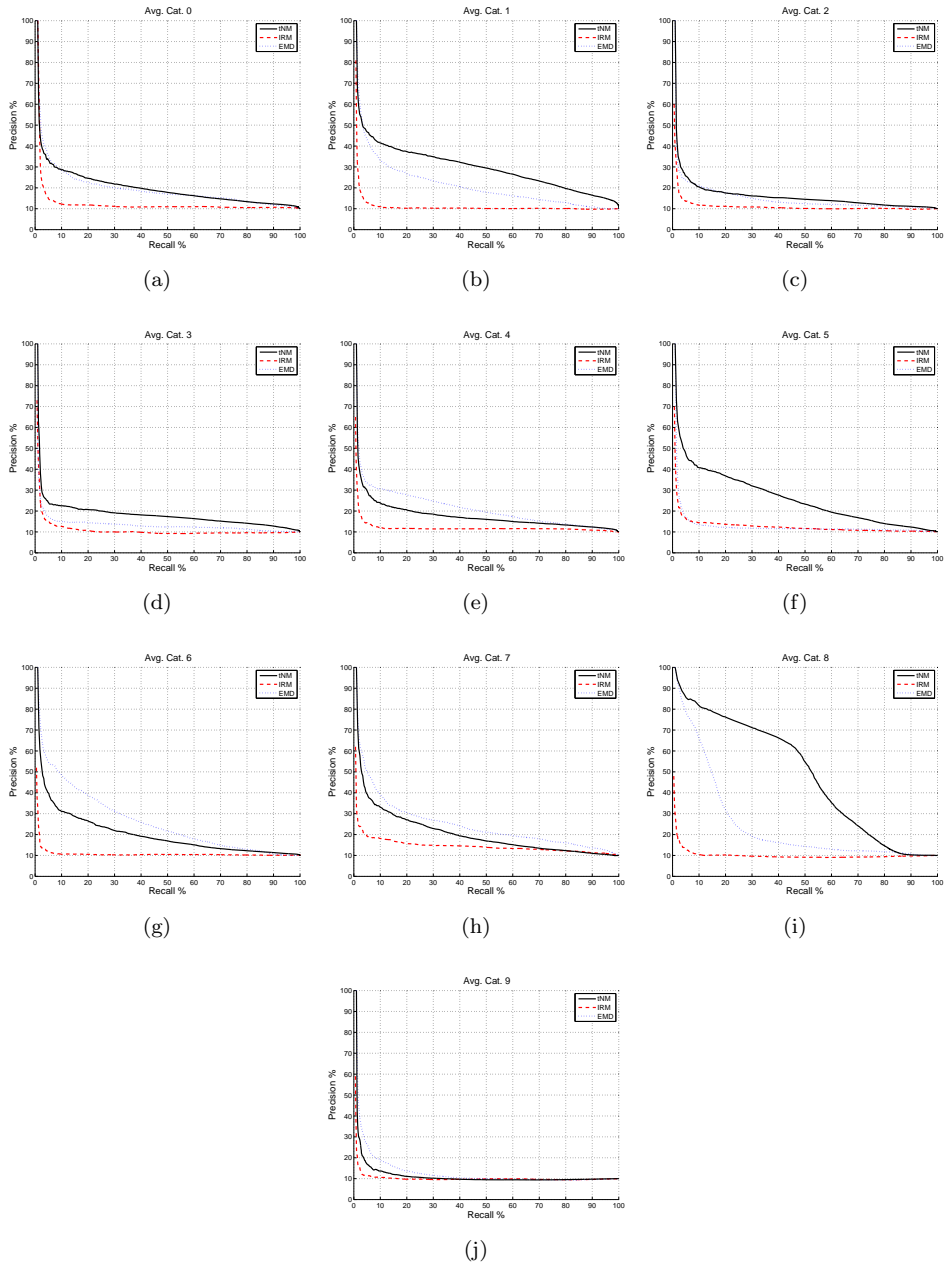FIGURE 7. Precision-recall for best results from each category.

FIGURE 8. Average precision-recall for each category.

Finally, a few comments on the differences in the methods presented here. The contrast in these approaches is due to the problem that the processes involved with image comparison are not well defined or understood. Specifically, the difference in the two approaches can be described as follows. The EMD and IRM similarity measures are based on determining the semantic similarity of points. In terms of image correspondence, the sets of points are based on image features contained in the signatures. The impetus of these methods is to measure the meaning (hence the term semantic) associated with the sets of points. To achieve this, the EMD and the IRM similarity measure take two different approaches. The EMD relies on distance functions (called ground distance [19]) to capture semantic similarity. On the other hand, the IRM similarity measure relies on the segmentation algorithm to isolate perceptual content of an image and uses soft matching of segments to correct for poor segmentations of the image perceptual content.

While the near set approach to quantifying the perceptual nearness of objects is not traditionally defined as signature-based, this framework can be applied to applications where the desired outcome is close to the human perception of nearness (as was the case in this article). The only requirement is that the problem must be able to be formulated in terms of sets of objects together with feature value vectors describing the objects. In order to understand the differences in the two approaches, it is important to distinguish between sets of points and perceptual objects. In the near set approach, perceptual objects are anything in the physical world with characteristics observable to the senses such that they can be measured and are knowable to the mind. Near set theory is used to assess similarity by extracting perceptually relevant information from objects grouped in classes based on object descriptions.

9. Conclusion

This article presented a comparison of the tNM with the popular EMD and IRM signature-based measures. The contribution of this paper is a new form of tNM for use in signature-based applications, where results were generated on a subset of the Corel image dataset other than those used in [21]. Results indicate that the tNM outperforms the IRM measure and is comparable to the EMD. Furthermore, the results presented here and those in [16] suggest that the tNM is a good alternative to the EMD in signature-based applications. Future work will consist of investigating whether different texture features will improve the ability of the tNM to assess similarity on images that contain predominately texture-based imagery, as well as, investigation into the claim that the tNM is as powerful an approach as the EMD.

References

[1] "knowledge": Oxford Dictionaries. Oxford University Press (April 2010)

[2] Peters, J.F., Wasilewski, P.: Foundations of near sets. Info. Sci. **179**(18) (2009) 3091–3109

[3] Pavel, M.: Fundamentals of Pattern Recognition. Marcel Dekker, Inc., NY (1993)

[4] Zeeman, E.C.: The topology of the brain and the visual perception. In Fort, K.M., ed.: Topoloy of 3-manifolds and selected topices. Prentice Hall, New Jersey (1965) 240–256

[5] Zeeman, E.C., Buneman, O.P.: Tolerance spaces and the brain. In Waddingto, C.H., ed.: Towards a theoretical biology. Aldine, London (1968) 140–151 Published in C.H. Waddington (Ed.), Towards a Theoretical Biology. The Origin of Life, Aldine Pub. Co.

[6] Naimpally, S.A.: Near and far. A centennial tribute to Frigyes Riesz. Siberian Electronic Mathematical Reports **6** (2009) A.1–A.10

[7] Naimpally, S.A., Warrack, B.D.: Proximity spaces. In: Cambridge Tract in Mathematics No. 59. Cambridge University Press, Cambridge, UK (1970)

[8] Ramanna, S., Meghdadi, A.H., Peters, J.F.: Nature-inspired framework for measuring image resemblance: A near rough set approach. Theoretical Computer Science **412**(42) (2011) 5926–5938 doi:10.1016/j.tcs.2011.05.044.

[9] Naimpally, S.A.: Proximity Approach to Problems in Topology and Analysis. Oldenburg Verlag, München (2009) ISBN 978-3-486-58917-7.

[10] Naimpally, S.A., Peters, J.F.: Topology with Applications.Topological Spaces via Near and Far. World Scientific, Singapore (2013) *In Press*.

[11] Henry, C.J.: Near Sets: Theory and Applications. PhD thesis (2010) Available at: `https://mspace.lib.umanitoba.ca/handle/1993/4267`.

[12] Wolski, M.: Perception and classification. A Note on near sets and rough sets. Fundamenta Informaticae **101** (2010) 143–155

[13] Peters, J.F.: Tolerance near sets and image correspondence. International Journal of Bio-Inspired Computation **1**(4) (2009) 239–245

[14] Peters, J.F.: Corrigenda and addenda: Tolerance near sets and image correspondence. International Journal of Bio-Inspired Computation **2**(5) (2010) 310–318

[15] Peters, J.F., Wasilewski, P.: Tolerance spaces: Origins, theoretical aspects and applications. Information Sciences **195**(0) (2012) 211–225

[16] Henry, C.J., Ramanna, S., Levy, D.: Quantifying nearness in visual spaces. Cybernetics and Systems Journal (2012) under review.

[17] Peters, J.F., Naimpally, S.A.: Applications of near sets. Notices of the American Mathematical Society **59**(4) (2012) 536–542

[18] Henry, C.J., Ramanna, S.: Parallel computation in finding near neighbourhoods. Lecture Notes in Computer Science **6954** (2011) 523–532

[19] Rubner, Y., Tomasi, C., Guibas, L.J.: The earth mover's distance as a metric for image retrieval. International Journal of Computer Vision **40**(2) (2000)

[20] Li, J., Wang, J.Z., Wiederhold, G.: IRM: Integrated region matching for image retrieval. In: Proceedings of the 8th ACM international conference on Multimedia. (2000) 147–156

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

[22] Peters, J.F.: Classification of objects by means of features. In: Proceedings of the IEEE Symposium Series on Foundations of Computational Intelligence (IEEE SCCI 2007). (2007) 1–8

[23] Peters, J.F.: Near sets. General theory about nearness of objects. Applied Mathematical Sciences **1**(53) (2007) 2609–2629

[24] Peters, J.F.: Classification of perceptual objects by means of features. International Journal of Information Technology & Intelligent Computing **3**(2) (2008) 1 – 35

[25] Duda, R., Hart, P., Stork, D.: Pattern Classification. 2nd edn. Wiley (2001)

[26] Sossinsky, A.B.: Tolerance space theory and some applications. Acta Applicandae Mathematicae: An International Survey Journal on Applying Mathematics and Mathematical Applications **5**(2) (1986) 137–167

[27] Henry, C.J.: Perceptually indiscernibility, rough sets, descriptively near sets, and image analysis. Transactions on Rough Sets **LNCS 7255** (2012) 41–121

[28] Hassanien, A.E., Abraham, A., Peters, J.F., Schaefer, G., Henry, C.: Rough sets and near sets in medical imaging: A review. IEEE Transactions on Information Technology in Biomedicine **13**(6) (2009) 955–968

[29] Henry, C.: Near set Evaluation And Recognition (NEAR) system. In Pal, S.K., Peters, J.F., eds.: Rough Fuzzy Analysis Foundations and Applications. CRC Press, Taylor & Francis Group (2010) 7–1 – 7–22 exe. availabe at: `http://wren.ee.umanitoba.ca`.

[30] Schroeder, M., Wright, M.: Tolerance and weak tolerance relations. Journal of Combinatorial Mathematics and Combinatorial Computing **11** (1992) 123–160

[31] Datta, D., Joshi, D., Li, J., Wang, J.Z.: Image retrieval: Ideas, influences, and trends of the new age. ACM Computing Surveys **40**(2) (2008) 5:1–5:60

[32] Hitchcock, F.L.: The distribution of a product from several sources to numerous localities. Journal of Mathematics and Physics **20** (1941) 224–230

[33] Rubner, Y., Tomasi, C., Guibas, L.J.: Code for the earth movers distance (emd), Stanford (1998) www.vision.caltech.edu/archive.html.

[34] Rubner, Y., Tomasi, C., Guibas, L.J.: A metric for distributions with applications to image databases. In: Proceedings of the 1998 IEEE International Conference on Computer Vision. (1998) 59–66

[35] Hillier, F.S., Lieberman, G.J.: Introduction to Mathematical Programming. McGraw-Hill, New York, NY (19980)

[36] Comaniciu, D.: Mean shift: A robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(5) (2002) 603–619

[37] Wang, J., Thiesson, B., Xu, Y., Cohen, M.: Image and video segmentation by anisotropic kernel mean shift. In: Proceedings of the European Conference on Computer Vision (ECCV'04). Volume 2. (2004) 238–249

[38] Christoudias, C., Georgescu, B., Meer, P.: Synergism in low level vision. In: Proceedings of the 16th International Conference on Pattern Recognition. Volume 4. (2002) 150–156

[39] Comaniciu, D., Ramesh, V., Meer, P.: The variable bandwidth mean shift and data-driven scale selection. In: Proceedings of the 8th IEEE International Conference on Computer Vision (ICCV 2001). Volume 1. (2001)

[40] Georgescu, B., Shimshoni, I., Meer, P.: Mean shift based clustering in high dimensions: A texture classification example. In: Proceedings of the 9th IEEE International Conference on Computer Vision (ICCV 2003). Volume 1. (2003) 456–463

[41] Garcia, V., Debreuva, E., Barland, M.: Fast k nearest neighbor search using GPU. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop. (2008) 1–6 Code URL: http://www.i3s.unice.fr/\textasciitildecreative/KNN/.

[42] Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(12) (2000) 1349–1380

[43] Choraś, R.S., Andrysiak, T., Choraś, M.: Integrated color, texture and shape information for content-based image retrieval. Pattern Analysis & Applications 10(4) (2007) 333–343

[44] Grigorova, A., De Natale, F.G.B., Dagli, C., Huang, T.S.: Content-based image retrieval by feature adaptation and relevance feedback. IEEE Transactions on Multimedia 9(6) (2007) 1183–1192

[45] Caicedo, J.C., González, F.A., Triana, E., Romero, E.: Design of Medical Image Database with Content-Based Retrieval Capabilities. Volume LNCS 4872 of Adavances in Image and Video Technology. Springer-Verlag, Berlin (2007)

[46] : The Corel Stock Photo Library (PC/MAC ver.). Corel Corporation, Ottawa, Ontario, Canada (1994)

Christopher J. Henry and Sheela Ramanna

Department of Applied Computer Science, University of Winnipeg, Canada, R3B 2E9

e-mail: `ch-henry@uwinnipeg.ca, s.ramanna@uwinnipeg.ca`