Preprint for Cybernetics and Systems 44(1) (2013): 38-56. DOI: 10.1080/01969722.2012.732797.

Quantifying Nearness in Visual Spaces

CHRISTOPHER J. HENRY, SHEELA RAMANNA, and DANIEL LEVY

Department of Applied Computer Science, University of Winnipeg, Canada

Cybernetic vision systems can be deployed in problem domains where the goal is to achieve results similar to those produced by humans. Fundamentally, these problems consist of evaluation of image content between sets of images. This paper contrasts two theoretical frameworks for image comparison, namely, the semantic similarity approach used in the Earth Mover's Distance (EMD) and the Integrated Region Matching (IRM) similarity measure, with the tolerance nearness measure (tNM) based on near set theory. The contribution of this article is a comparison of the image similarity measures EMD, IRM, and tNM, as well as a signature-based approach to calculating the tolerance nearness measure.

KEYWORDS: tolerance space, vision system, tolerance nearness measure, earth mover's distance, integrated region matching

INTRODUCTION

The problem considered in this article is one of synthesizing human perception of nearness for use in cybernetic vision systems. Many areas of research into theoretical and practical applications of the human concepts of similarity, proximity, and nearness, have, at their heart (either intentionally or otherwise), a framework for evaluating the similarity of objects, a task that we perform effortlessly and without thought on a daily basis. Examples include cybernetic computer vision systems, image segmentation systems, image object recognition systems, and content-based image retrieval systems. The focus of this article is on finding similarities between digital images, an approach that can be used in the design of vision systems. Specifically, the near set approach to measuring the resemblance between images extracted from video sequences reported in (Henry and Peters

Address correspondence to Christopher Henry, Department of Applied Computer Science, University of Winnipeg, 515 Portage Avenue, Winnipeg, Manitoba, Canada R3B 2E9. E-mail: ch.henry@uwinnipeg.ca

2011) is compared to the popular Earth Mover's Distance (EMD) (Rubner et al. 2000) and the Integrated Region Matching (IRM) similarity measuring (Li et al. 2000; Wang et al. 2001), which are traditionally not associated with vision systems. However, all these methods were created in an effort to find practical applications that encode our ability to assess similarity.

(Wang et al. 2001) captures the essence of the problem when they state that it is straightforward to measure the distance between two points, yet, it is difficult to quantify the distance between sets of points, especially when these sets of points are, in some fashion, tied to our perception of the content captured by the image. In other words, the desired output of any vision system is fundamentally tied to our innate ability to rapidly evaluate and quantify the similarity of objects in our environment. The comparison of methods presented in this paper represents an advance between approaches that measure the semantic similarity between sets of points, as is the case with the Earth Mover's Distance (EMD) (Rubner et al. 2000) and the Integrated Region Matching (IRM) similarity measuring (Li et al. 2000; Wang et al. 2001), and those based on recently introduced near set theory (Peters 2007a,b; Peters and Wasilewski 2009; Ramanna et al. 2011), such as the tolerance nearness measure (tNM) (Henry and Peters 2011; Henry 2010b) that assess the perceptual nearness of objects. These three approaches were inspired by ideas ranging from the desire to reduce the amount of work involved in moving piles of dirt to a collection of holes; a desire to quantify the semantic "closeness" of two images; and observations about the tolerance introduced by our senses and an intuitive approach to assessing nearness of objects. This article is organized as follows: we first start with a discussion of related works, followed by an introduction to tolerance near sets. This discussion is followed by an in-depth look at signature-based methods and our tolerance nearness measure. The last part of the paper is devoted to implementation details and an analysis of the results.

RELATED WORKS

The near set approach reported here is based on tolerance spaces and is related to both physical and visual spaces, which are the domain of visual systems. As defined by (Wagner 2006), a physical space is the space revealed by instrumentation and is independent of the observer, while a visual space is a non-objective interpretation by the observer of the physical space based on the perception of external stimuli. For instance, it is impossible to determine the accuracy of a person's judgement of the properties of a physical room, whereas these values can be obtained exactly through measurement. In the former case, for example, it could be that some shadows caused the observer to misjudge the actual shape of the room. Wagner's definition of a visual space describes the environment of a visual system, which is an artificial approach to mimicking the human visual system. Generally, these systems consist of a sensing device (such as a camera) that generates an image, or stream of images, as well as a processing unit that interprets the images and makes decisions. As a result, a visual system operates in a visual space since the judgements are based only on the output of the sensors.

Notice the underlying theme in the above discussion, *i.e.* the problem domains in which visual systems operate are based on an assessment of similarity, proximity, or nearness of perceived objects, where these systems are designed to mimic human decision processes based on quantified feature values obtained from sensors. Naturally, these systems require a theoretical framework for making decisions. It is these frameworks, and their application, that are the focus of work in many different fields. For example, the near set approach used here, introduced by Peters (Peters 2007a,b; Peters and Wasilewski 2009) (see also (Wolski 2010, 2011; Abd El-Monsef et al. 2010)) establishes a formal basis for identifying, comparing, and measuring resemblances of objects based on their descriptions, *i.e.* based on the features that describe the objects. Other examples include the work by Horst Herrlich (Herrlich 1974), that aims to create a unifying concept of nearness that encompasses many other categories based on the simple idea of the nearness of collections of sets. Namely, his work covers topological $\mathrm{R_O}$-spaces, continuous maps, uniform spaces and uniformly continuous maps, proximity spaces (see also (Naimpally 2009; Naimpally and Warrack 1970)) and δ-maps, and contiguity maps.

Also inherent to this discussion is the idea of perception. The term *perception*[1] appears in the literature in many different places with respect to images and visual systems. For instance, the term is often used for demonstrating that the performance of methods are similar to results obtained

[1] See, (Martin and Gordon 2001) for an interesting discussion on the evolution of perception

by human subjects (as in (Montag and Fairchild 1997)), or it is used when the system is trained from data generated by human subjects (as in (El-Naqa et al. 2004)). 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 (Rahman et al. 2007; Martinez et al. 2005). For instance, (Martinez et al. 2005) 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, users might want to search for a tall tree they remembered seeing on one of their visits to a virtual city.

Other interpretations of *perception* are tightly coupled to psychophysics, *i.e.* perception based on the relationship between stimuli and sensation (Bruce et al. 1996). For instance, the idea of tolerance first surfaced in Poincaré's work in 1905 (Poincaré 1905; Henry 2010b) in which he reflects on psychophysics experiments performed by Ernst Weber in 1834, and Gustav Fechner's insight in 1850 (Sossinsky 1986; Benjamin 2007; Hergenhahn 2009; Fechner 1966). More recently, (Papathomas et al. 1997) introduces a texture perception model. The texture perception model uses the antagonistic view of the human visual system 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.* (Mojsilovic et al. 2002), where it is suggested that humans view or recall an image by its dominant colours only, and areas containing small, non-dominant colours are averaged by the human visual system. Other examples of the term perception defined in the context of psychophysics have also been given (Balakrishnan et al. 2005; Qamra et al. 2005; Wang et al. 2004b; Dempere-Marco et al. 2002; Kuo and Johnson 2002; Wandell et al. 2002; Wilson et al. 1997).

Perception as explained by psychologists (Hoogs et al. 2003; Bourbakis 2002) 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, (Bourbakis 2002) 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.

As was mentioned in the introduction, a vision system is one that that mimics the power and capability of the human sense of sight (*i.e.* the ability to detect light) combined with some type of cognition, perception, or interpretation of the stimulus. We propose that the approach to measuring the similarities of images presented in this article could be useful in the design of a visual system. While a complete survey of vision systems is outside the scope of this article, the following examples are presented to give an idea as to the various types of vision systems. (Bakhtari and Benhabib 2007) present a vision system with the goal to position multiple cameras to identify and track multiple objects of interest in dynamic multiobject environments. (Hussmann and Liepert 2009) use 3-D time of flight (rather than stereo vision) to control a robot in a simulation of loading a container ship. The visual system generates range data to the objects that need to be loaded onto a ship, and performs segmentation of an image generated from range data to identify the centre of gravity and the rotation angle (information necessary to grab the simulated containers). Finally, another example of a vision system is the CogV system presented in (Zhang and Tay 2009, 2011) which mimics saccade and vergence movements in a binocular camera system to identify objects of interest in the field of view.

TOLERANCE NEAR SETS

Nearness is an intuitive concept useful in comparing the descriptions of objects encountered in our daily lives. At a young age, we become adept at detecting similarities of objects in our environment and quickly assessing degrees of similarity. In fact, our day-to-day conversations are full of adverbs (e.g., closely, nearly), adjectives (e.g., alike, almost, similar) and nouns (e.g., affinities) used in expressing the nearness of “things” with common characteristics.

Near sets are disjoint sets that resemble each other. All sets in near set theory consist of perceptual objects, which are anything in the physical world with characteristics observable to the senses such that they can be measured and are knowable to the mind. In the context of near set theory, objects in our visual field are always presented with respect to the selected probe functions.

A probe function is a real-valued function representing a feature of a perceptual object (Peters 2007a). This is in keeping with the approach to pattern recognition suggested by M. Pavel (Pavel 1993) where the features of an object are quantified by probe functions. In other words, probe functions are used to measure characteristics of visual objects and similarities among perceptual objects.

A perceptual system is a set of perceptual objects, together with a set of probe functions, *i*.e. 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 $\phi : O \to \mathbb{R}$ (Peters and Wasilewski 2009). The notion of a perceptual system admits a wide variety of different interpretations that result from the selection of sample perceptual objects contained in a particular sample space O. Two examples of perceptual systems are: a set of images together with a set of image processing probe functions, or a set of results from a web query together with some measures (probe functions) indicating, *e.g.*, relevancy or distance (*i.e.* geographical or conceptual distance) between web sources. 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

$$\boldsymbol{\phi}_{\mathcal{B}}(x) = (\phi_1(x), \phi_2(x), \ldots, \phi_i(x), \ldots, \phi_l(x)), \tag{1}$$

where l is the length of the vector $\boldsymbol{\phi}_{\mathcal{B}}$, and each $\phi_i(x)$ in $\boldsymbol{\phi}_{\mathcal{B}}(x)$ 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 (Duda et al. 2001). The description of an object can be considered a point in an l-dimensional Euclidean space $\mathbb{R}^l$ called a feature space. Thus, the relationship between objects is discovered in a feature space that is determined by the probe functions in $\mathcal{B}$.

Eq. (1) plays a central role in the perceptual indiscernibility relation and the tolerance relation, which are used in the definition of near[2] and tolerance near sets respectively. The tolerance relation

[2] The results presented here are based on tolerance near sets, and, consequently, a discussion on near set theory is

is defined within the context of a tolerance space. Let O be a set of sample perceptual objects, and let ξ be a binary relation (called a tolerance relation) on X ($\xi \subset X \times X$) 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 $\langle X, \xi \rangle$. Considering the tolerance space definition, a specific tolerance relation (Peters 2009, 2010) (see (Hassanien et al. 2009; Henry 2010a) for applications) is given as follows. 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 $\cong_{\mathcal{B},\epsilon}$ is defined by:

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

where $\| \cdot \|_2$ 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 perceptual tolerance relation gives two very different and useful classes due to its lack of transitivity. Let $\langle O, \mathbb{F} \rangle$ be a perceptual system and let $x \in O$. For a set $\mathcal{B} \subseteq \mathbb{F}$ and $\varepsilon \in \mathbb{R}$, a neighbourhood is defined as $N(x) = \{y \in O : x \cong_{\mathcal{B},\epsilon} y\}$. In contrast, all the pairs of objects within a pre-class must satisfy the tolerance relation. 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 $x \cong_{\mathcal{B},\epsilon} y$ for any pair $x, y \in X$. A maximal pre-class with respect to inclusion is called a tolerance class. 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, tolerance near sets can be defined by way of the tolerance nearness relation (Peters 2009, 2010). 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 (*i.e.* X and Y satisfy the tolerance nearness relation) within the perceptual system $\langle O, \mathbb{F} \rangle$ ($X \underset{\cong}{\bowtie}_{\mathbb{F}} Y$) iff there exists $x \in X$ and $y \in Y$ and there is $\mathcal{B} \subseteq \mathbb{F}$ such that $x \cong_{\mathcal{B},\epsilon} y$. Using the tolerance nearness relation, tolerance near sets can be defined as follows (Peters 2009, 2010). Let $\langle O, \mathbb{F} \rangle$ be a perceptual system and let $\varepsilon \in \mathbb{R}, \mathcal{B} \subseteq \mathbb{F}$. Further, let $X, Y \subseteq O$, denote disjoint

outside the scope of this article. See (Henry 2010b) for comprehensive review of near sets (and tolerance near sets).

sets with coverings determined by the tolerance relation $\cong_{\mathcal{B},\epsilon}$, and let $H_{\cong_{\mathcal{B},\varepsilon}}(X), H_{\cong_{\mathcal{B},\varepsilon}}(Y)$ denote the set of tolerance classes for X, Y, respectively. Sets X, Y are tolerance near sets iff there are tolerance classes $A \in H_{\cong_{\mathcal{B},\varepsilon}}(X), B \in H_{\cong_{\mathcal{B},\varepsilon}}(Y)$ such that $A \underset{=}{\bowtie}_{\mathbb{F}} B$.

SIGNATURE-BASED MEASURES

This section presents signature-based methods investigated in this article. Generally, signatures are mathematical descriptions of an image (Datta et al. 2008), where specific signature details are dependent on the application. We define a signature as a feature vector associated with a set of pixels 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 the Implementation Section). Then, the dominant RGB colour and pixel count constituted the signature for each region.

EARTH MOVER'S DISTANCE

The EMD was introduced by Rubner in (Rubner et al. 2000) and is also known as Mallows distance when applied to probability frequencies (Datta et al. 2008). 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, subject to constraints, where signatures from the respective images are cast as consumers and suppliers.

Specifically, for two images, let $R_1 = \{(\mathbf{r}_1, w_{\mathbf{r}_1}), \ldots, (\mathbf{r}_m, w_{\mathbf{r}_m})\}$ (resp. $R_2 = \{(\mathbf{r}'_1, w_{\mathbf{r}'_1}), \ldots,$ $(\mathbf{r}_n, w_{\mathbf{r}'_n})\}$) be the first (second) signature, where $\mathbf{r}_i$ ($\mathbf{r}'_j$) is the cluster representative (called region descriptor in (Wang et al. 2001)) and $w_{\mathbf{r}_i}$ ($w_{\mathbf{r}'_j}$) is the weight of each cluster (we used the cluster area percentage scheme (Wang et al. 2001)). Furthermore, let $d(\mathbf{r}_i, \mathbf{r}'_j)$ be the distance between region cluster representatives. Then, to calculate the EMD, it is necessary to find a flow $F =$

$\{f_{i,j} : i = 1, \ldots, m; j = 1, \ldots, n\}$ that minimizes

$$\text{WORK}(R_1, R_2, F) = \sum_{i=1}^{m} \sum_{j=1}^{n} d(\mathbf{r}_i, \mathbf{r}'_j) f_{i,j},$$

subject to constraints (reported in (Rubner et al. 2000)) on direction and amount of supplies sent by clusters, and on the amount of supplies received by clusters. Additionally, clusters are also required to move the maximum amount of supplies possible between clusters. Once the optimal flow F has been calculated, the EMD is defined as

$$\text{EMD}(R_1, R_2) = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} d(\mathbf{r}_i, \mathbf{r}'_j) f_{i,j}}{\sum_{i=1}^{m} \sum_{j=1}^{n}}.$$

INTEGRATED REGION MATCHING SIMILARITY MEASURE

The IRM similarity measure is a soft matching approach to measuring the similarity of images (Li et al. 2000; Wang et al. 2001). 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) (Datta et al. 2008). Here, it is important to differentiate between segments or regions, and perceptual concepts within an image. Let us define segments and 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.

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 (Wang et al. 2001). Let S represent a significance matrix indicating the importance between matching region $\mathbf{r}_i$ with region $\mathbf{r}'_j$. Then, the IRM similarity measure is defined as

$$d(R_1, R_2) = \sum_{i=1}^{m} \sum_{j=1}^{n} s_{i,j} d(\mathbf{r}_i, \mathbf{r}'_j).$$

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.*, (Wang et al. 2001)). The algorithm used to populate S is given in (Wang et al. 2001).

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.* (Hassanien et al. 2009; Henry 2010b)). 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 is defined as follows. Let $\langle O, \mathbb{F} \rangle$ be a perceptual system, with $\varepsilon \in \mathbb{R}$, and $\mathcal{B} \subseteq \mathbb{F}$. Furthermore, let X and Y be two disjoint sets and let $Z = X \cup Y$. Then a tolerance nearness measure between two sets is given by

$$tNM_{\cong_{\mathcal{B},\varepsilon}}(X, Y) = \\ 1 - \left(\sum_{C \in H_{\cong_{\mathcal{B},\varepsilon}}(Z)} |C| \right)^{-1} \cdot \sum_{C \in H_{\cong_{\mathcal{B},\varepsilon}}(Z)} |C| \frac{\min(|C \cap X|, |[C \cap Y|)}{\max(|C \cap X|, |C \cap Y|)}. \quad (2)$$

Traditionally, the tNM has been used in measuring nearness in problem domains that generate many objects for comparison. For example, in (Henry 2010b), 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. 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 signatures for two similar regions from the respective images, or small groups of signatures from each region Thus, in order to provide a basis for comparison, we needed to adapt the tolerance near set approach to take into consideration the fact that we are matching regions, with few or even one signature, rather than a large set of objects. As a result, only the signature values were used to form tolerance classes, and the cardinality $|C|$ in Eq. 2 was replaced by the total region count of all signatures in C.

IMPLEMENTATION

All the signatures used to generate our results were created by an adaptive mean shift segmentation algorithm. The mean shift algorithm, introduced in (Comaniciu 2002), 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) (Wang et al. 2004a). 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 (Duda et al. 2001). The mean shift segmentation algorithm used in this article were created using our own modification of the EDISON system (Christoudias et al. 2002), 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 (Comaniciu et al. 2001). 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, we used the approach reported in (Georgescu et al. 2003) (also described in (Duda et al. 2001)), where for each pixel in the image the bandwidth parameter is the distance to its k^{th} 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[3].

Kernel density estimators that vary the bandwidth parameter based on the k^{th}-nearest neighbour are called balloon density estimators (Comaniciu et al. 2001). 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 k^{th}-nearest neighbour. Clearly, the success of this approach relies on a fast solution to the k-nearest neighbour search problem. To solve this problem we used the KNN CUBLAS GPU implementation reported in (Garcia et al. 2008). Note, there are some disadvantages of using the balloon density approach. However, our aim was to relieve the burden of finding a globally acceptable bandwidth, or having to select a bandwidth parameter for each image. The k^{th}-nearest neighbour solved both these problems, while providing good segmentations.

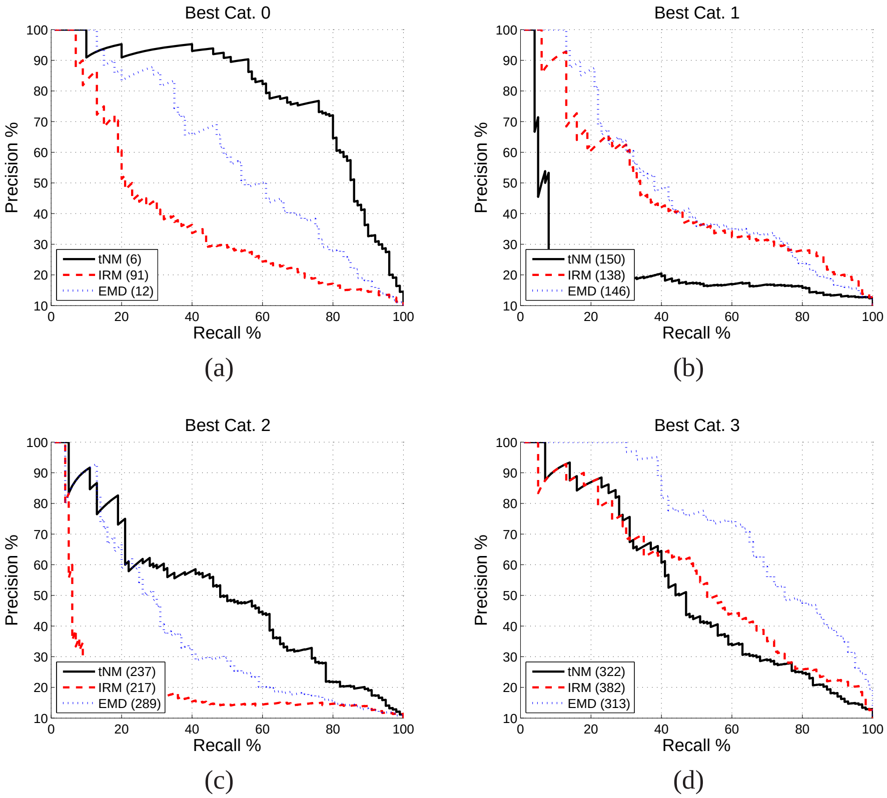
RESULTS AND DISCUSSION

This section presents the results of comparing the signature-based methods presented above. Specifically, the results were generated with the SIMPLIcity image database (Wang et al. 2001; Li and Wang 2003), a database of images containing 10 categories with 100 images in each category. The categories are varied with different objects and scenes, and images in different categories can also resemble each other. These results are intended to demonstrate using each measure to assess similarity of images for use in cybernetic vision systems. Consequently, the results are presented using precision-recall plots (Yates-Baeza and Ribeiro-Neto 1999) that show the similarity of a single image to all other images in a database (the idea being images from the same category should

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

generate the smallest measure values).

The results are presented in Fig. 1 - 3, where each plot represents best results for each category. We defined the best result as the query image that generated the most precision values over 85%. Notice, for the most part, the results of the tNM are comparable to the EMD and the IRM similarity measure; with the tNM outperforming the others for categories 0, and 2; underperforming in categories 1, 4, and 5; and generating similar results for the rest.

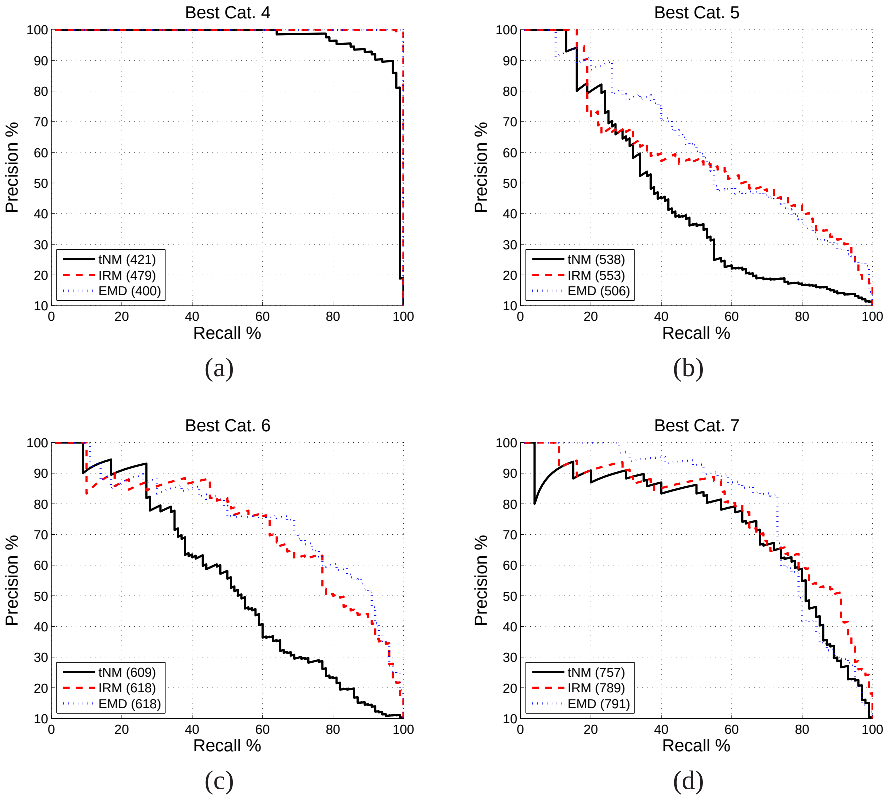


(a) (b)

(c) (d)

FIGURE 1 Precision-recall for best results: Categories 0-3.

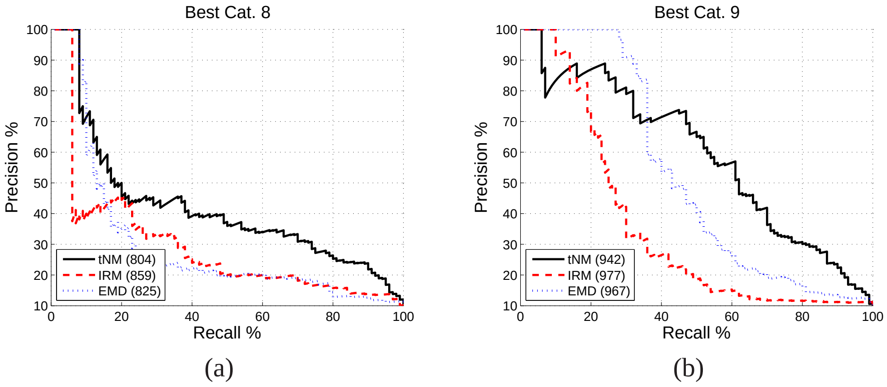
Observe that while the tNM performs quite well for category 4, it still is outperformed by the other measures. This is due to the nature of the images and the signatures they generate. In particular, the images in category 4 are all drawings of dinosaurs on a light background, where the background was easily segmented from the dinosaur. This means the background pixels in these images are distributed among a small number of signatures. The result is small tolerance classes, causing great variation in the tNM. For example, consider two scenarios for a pair of dinosaur images: first, each background is represented by a single signature, and, second, one



(a) (b)

(c) (d)

FIGURE 2 Precision-recall for best results: Categories 4-7.



(a) (b)

FIGURE 3 Precision-recall for best results: Categories 8-9.

background is represented by a single signature and another with two signatures. These cases will generate proper fractions of 1, and 0.5 respectively (with respect to the calculation of the tNM in Eq. 2). Since each proper fraction is weighted by the number of pixels in a region, these proper fractions will significantly affect the outcome of the tNM due to the size of the background in each image. Moreover, if, from the previous example, the image with two signatures representing the background forms a tolerance class with just two or three other signatures (giving a proper fraction of either 1 or 2/3) from an image in another category, it is very likely that this non-dinosaur will be ranked higher than the dinosaur image with one signature representing the background. The solution to this problem is to use subimages instead of signatures (as reported in (Henry 2010b)), in which case it is possible to achieve precision-recall values like those given in Fig. 4.

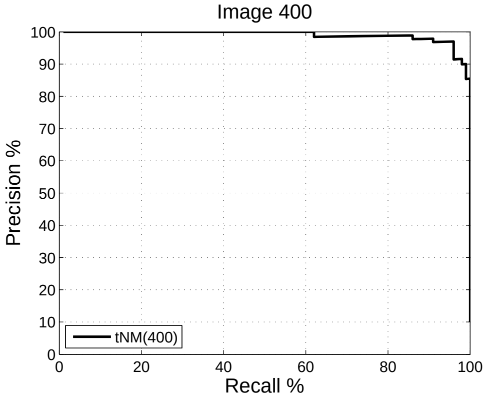


FIGURE 4 Results generated using subimages instead of signatures for images from category 4.

The tNM was also outperformed in categories 1 and 5. Nevertheless, the retrieved images for these categories are still perceptually near each other with respect to the selected probe functions, *i.e.* the retrieved images are similar with respect to colour content. This can be seen by Fig. 5 & 6, where Fig. 5 contains the best image retrieval results for category 1. The three rows in this figure contain the images with the smallest distance (from left to right) with respect to the best query image for each measure, where the first image in each row is the query image (and, hence, had the smallest distance). In turn, Fig. 6 contains colour cloud plots detailing pixel image colour in the RGB colour space and were generated by plotting each of the colours in an image, where the precise location of a point is determined by a normal distribution with a standard deviation proportional to the number of pixels belonging to a specific colour. Consequently, colours that occur more often in the image will occupy a larger volume in the plot than colours that do not. Comparing Fig. 6 & 7, it is easy to see that, while the images in Fig. 5a-5e do not all belong to the same category, they are similar to each other with respect to pixel colour.

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

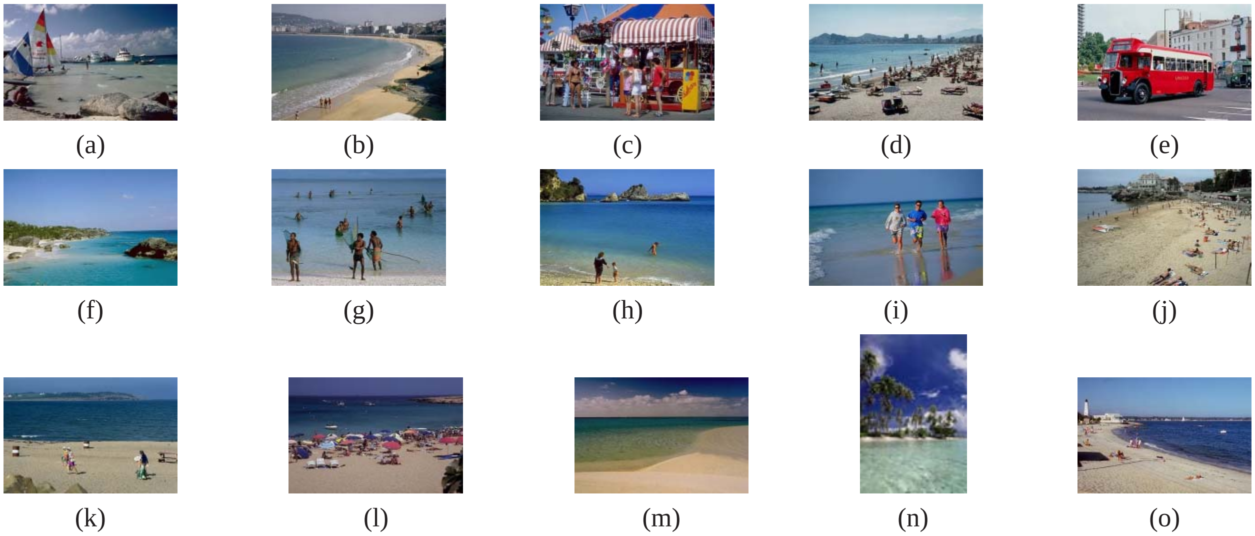
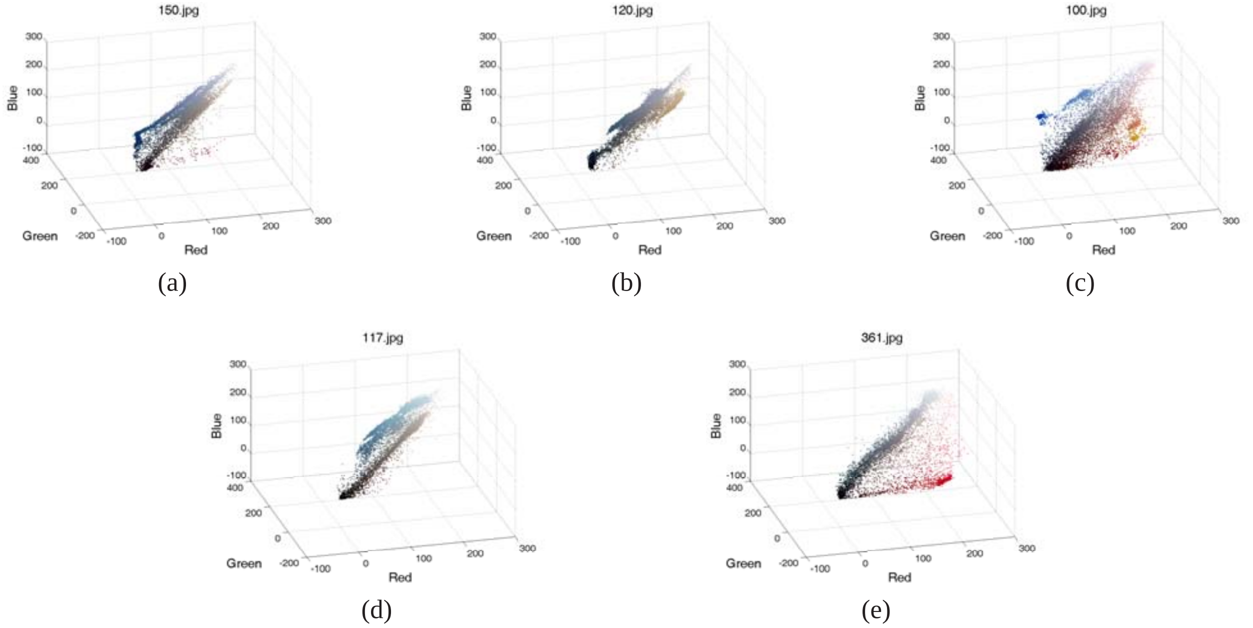

(a) (b) (c) (d) (e)
(f) (g) (h) (i) (j)
(k) (l) (m) (n) (o)

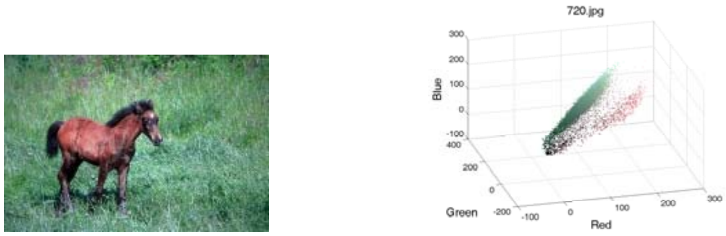
FIGURE 5 Top query results from category 1 for each measure: Rows from top to bottom respectively denote, tNM, IRM, EMD, the first image in each row is the best query image, and the remaining images are the most similar to the query based on measure values.



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

FIGURE 6 Colour cloud plots of Fig. 5a-5e showing image pixel colour in RGB colour space.

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



(a) (b)

FIGURE 7 Example of colour cloud plots that are not similar to those in Fig. 6.

measure take two different approaches. The EMD relies on distance functions (called ground distance Rubner et al. (2000)) 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.

CONCLUSION

This article presented a comparison between approaches evaluating semantic similarity of sets of points, and perceptual nearness of objects. Particularly, the contribution of this article is a comparison of the image similarity measures EMD, IRM, and tNM for use in cybernetic vision systems, as well as a signature-based approach to calculating the tolerance nearness measure. Results indicate there was no single approach that clearly outperformed all the others. The work presented here is a first step toward claiming the tNM is an approach as powerful as the well known EMD and IRM similarity measure. Future work will consist of further comparisons of these methods with the addition of texture and edge based features, rather than only the colours features that were used in this article, and investigations into the affect of using an adaptive mean shift algorithm based on sample point estimators rather than balloon estimators.

ACKNOWLEDGEMENTS

This research has been supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) grant 194376, the University of Winnipeg Research Start-Up Grant, and the University of Winnipeg Major Research Grant.

REFERENCES

Abd El-Monsef, M. E., A. M. Kozae, and M. J. Iqelan. 2010. Near approximations in topological spaces. *International Journal of Mathematical Analysis* 4(6):279–290.

Bakhtari, A., and B. Benhabib. 2007. An active vision system of multitarget surveillance in dynamic environments. *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics* 37(1):190–198.

Balakrishnan, N., K. Hariharakrishnan, and D. Schonfeld. 2005. 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* 27(9):1367–1378.

Benjamin, L. T., Jr. 2007. *A brief history of modern psychology*. Malden, MA: Blackwell Publishing.

Bourbakis, N. G. 2002. Emulating human visual perception for measuring difference in images using an spn graph approach. *IEEE Transactions on Systems, Man, and Cybernetics, Part B* 32(2):191–201.

Bruce, V., P. R. Green, and M. A. Georgeson. 1996. *Visual perception: physiology, psychology, and ecology*. Hove, East Sussex, UK: Psychology Press.

Christoudias, C., B. Georgescu, and P. Meer. 2002. Synergism in low level vision. In *Proceedings of the 16th international conference on pattern recognition*, vol. 4, 150–156.

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

Comaniciu, D., V. Ramesh, and P. Meer. 2001. The variable bandwidth mean shift and data-driven scale selection. In *Proceedings of the 8th ieee international conference on computer vision (iccv 2001)*, vol. 1, 438–445.

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

Dempere-Marco, L., Hu Xiao-Peng, S. L. S. MacDonald, S. M. Ellis, D. M. Hansell, and Guang-Zhong Yang. 2002. The use of visual search for knowledge gathering in image decision support. *IEEE Transactions on Medical Imaging* 21(7):741–754.

Duda, R.O., P.E. Hart, and D.G. Stork. 2001. *Pattern classification*. 2nd ed. Wiley.

El-Naqa, I., Yongyi Yang, N. P. Galatsanos, R. M. Nishikawa, and M. N. Wernick. 2004. A similarity learning approach to content-based image retrieval: application to digital mammography. *IEEE Transactions on Medical Imaging* 23(10):1233–1244.

Fechner, G. T. 1966. *Elements of psychophysics, vol. i*. London, UK: Hold, Rinehart & Winston. H. E. Adler's trans. of Elemente der Psychophysik, 1860.

Garcia, V., E. Debreuva, and M. Barland. 2008. Fast k nearest neighbor search using GPU. In *Ieee computer society conference on computer vision and pattern recognition workshop*, 1–6. Code URL: `http://www.i3s.unice.fr/~creative/KNN/`.

Georgescu, B., I. Shimshoni, and P. Meer. 2003. Mean shift based clustering in high dimensions: A texture classification example. In *Proceedings of the 9th ieee international conference on computer vision (iccv 2003)*, vol. 1, 456–463.

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

Henry, C. 2010a. Near set Evaluation And Recognition (NEAR) system. In *Rough fuzzy analysis foundations and applications*, ed. S. K. Pal and J. F. Peters, 7–1 – 7–22. CRC Press, Taylor & Francis Group. exe. availabe at: `http://wren.ee.umanitoba.ca`.

Henry, C. J. 2010b. Near Sets: Theory and Applications. Ph.D. thesis. Available at: `https://mspace.lib.umanitoba.ca/handle/1993/4267`.

Henry, C. J., and J. F. Peters. 2011. Neighbourhood-based vision systems. *Cybernetics and Systems* 42(1): 33–44.

Hergenhahn, B. R. 2009. *An introduction to the history of psychology*. Belmont, CA: Wadsworth Publishing.

Herrlich, H. 1974. A concept of nearness. *General Topology and its Applications* 5:191–212.

Hoogs, A., R. Collins, R. Kaucic, and J. Mundy. 2003. A common set of perceptual observables for grouping, figure-ground discrimination, and texture classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25(4):458–474.

Hussmann, S., and T. Liepert. 2009. Three-dimensional tof robot vision system. *IEEE Transactions on Instrumentation and Measurement* 58(1):141–146.

Kuo, S., and J. D. Johnson. 2002. Spatial noise shaping based on human visual sensitivity and its application to image coding. *IEEE Transactions on Image Processing* 11(5):509–517.

Li, J., and J. Z. Wang. 2003. Automatic linguistic indexing of pictures by a statistical modeling approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25(9):1075–1088.

Li, J., J. Z. Wang, and G. Wiederhold. 2000. Irm: Integrated region matching for image retrieval. In *Proceedings of the 8 acm international conference on multimedia*, 147–156.

Martin, C. C., and R. Gordon. 2001. The evolution of perception. *Cybernetics and Systems* 32(3-4):393–409.

Martinez, J. I., A. F. G. Skarmeta, and J. B. Gimeno. 2005. Fuzzy approach to the intelligent management of virtual spaces. *IEEE Transactions on Systems, Man, and Cybernetics, Part B* 36(3):494–508.

Mojsilovic, A., H. Hu, and E. Soljanin. 2002. Extraction of perceptually important colors and similarity measurement for image matching, retrieval and analysis. *IEEE Transactions on Image Processing* 11(11): 1238–1248.

Montag, E. D., and M. D Fairchild. 1997. Pyschophysical evaluation of gamut mapping techniques using simple rendered images and artificial gamut boundaries. *IEEE Transactions on Image Processing* 6(7): 977–989.

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

Naimpally, S. A., and B. D. Warrack. 1970. Proximity spaces. In *Cambridge tract in mathematics no. 59*. Cambridge, UK: Cambridge University Press.

Papathomas, T. V., R. S. Kashi, and A. Gorea. 1997. A human vision based computational model for chromatic texture segregation. *IEEE Transactions on Systems, Man, and Cybernetics, Part B* 27(3): 428–440.

Pavel, M. 1993. *Fundamentals of pattern recognition*. NY: Marcel Dekker, Inc.

Peters, J. F. 2007a. Near sets. General theory about nearness of objects. *Applied Mathematical Sciences*

1(53):2609–2629.

———. 2007b. Near sets. Special theory about nearness of objects. *Fundamenta Informaticae* 75(1-4): 407–433.

———. 2009. Tolerance near sets and image correspondence. *International Journal of Bio-Inspired Computation* 1(4):239–245.

———. 2010. Corrigenda and addenda: Tolerance near sets and image correspondence. *International Journal of Bio-Inspired Computation* 2(5):310–318.

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

Poincaré, H. 1905. *Science and hypothesis*. Brock University: The Mead Project. L. G. Ward's translation.

Qamra, A., Y. Meng, and E. Y. Chang. 2005. Enhanced perceptual distance functions and indexing for image replica recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(3):379–391.

Rahman, M., P. Bhattacharya, and B. C. Desai. 2007. A framework for medical image retrieval using machine learning and statistical similarity matching techniques with relevance feedback. *IEEE Transactions on Information Technology in Biomedicine* 11(1):58–69.

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

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

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

Wagner, M. 2006. *The geometries of visual space*. Mahwah, New Jersey, USA: Lawrence Erlbaum Associates, Inc.

Wandell, B. A., A. El Gamal, and B. Girod. 2002. Common principles of image acquisition systems and biological vision. *Proceedings of the IEEE* 90(1):5–17.

Wang, J., B. Thiesson, Y. Xu, and M. Cohen. 2004a. Image and video segmentation by anisotropic kernel mean shift. In *Proceedings of the european conference on computer vision (eccv'04)*, vol. 2, 238–249.

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

Wang, Z., A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. 2004b. Image quality assesment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* 13(4):600–612.

Wilson, T. A., S. K. Rogers, and M Kabrisky. 1997. Perceptual-based image fusion for hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing* 35(4):1007–1017.

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

———. 2011. Gauges, pregauges and completions: Some theoretical aspects of near and rough set approaches to data. In *Proceedings of the 6 international conference on rough sets and knowledge technology (rskt11)*, vol. LNCS 6954, 559–568. Springer.

Yates-Baeza, R., and B. Ribeiro-Neto. 1999. *Modern information retrieval*. New York: ACM Pres/Pearson Addison Wesley.

Zhang, X., and A. L. P. Tay. 2009. A physical system for binocular vision through saccade generation and vergence control. *Cybernetics and Systems* 40:549–568.

———. 2011. A binocular vision system with attentive saccade and spatial variant vergence control. *Cybernetics and Systems Analysis* 42(1):45–63.